PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, A61K 38/17, C12Q 1/68

A2

(11) International Publication Number: WO 00/15799

(43) International Publication Date: 23 March 2000 (23.03.00)

(21) International Application Number: PCT/US99/21688

(22) International Filing Date: 17 September 1999 (17.09.99)

(30) Priority Data:

60/155,246	17 September 1998 (17.09.98)	US
09/158,720	22 September 1998 (22.09.98)	US
Not furnished	22 September 1998 (22.09.98)	US
60/069,391	4 November 1998 (04.11.98)	US
60/128,660	8 April 1999 (08.04.99)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	Not furnished (CIP)
Filed on	17 September 1998 (17.09.98)
US	09/158,720 (CIP)
Filed on	22 September 1998 (22.09.98)
US	Not furnished (CIP)
Filed on	22 September 1998 (22.09.98)
US	09/186,815 (CIP)
Filed on	4 November 1998 (04.11.98)
US	09/156,039 (CIP)
Filed on	17 September 1998 (17.09.98)
US	Not furnished (CIP)
Filed on	4 November 1998 (04.11.98)
US	60/128,660 (CIP)
Filed on	8 April 1999 (08.04.99)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). GORGONE, Gina, A. [US/US]; 1253 Pinecrest Drive, Boulder Creek, CA 95006 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). YANG, Junming [CN/US]; 7136 Clarendon Street, San Jose, CA 95129 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

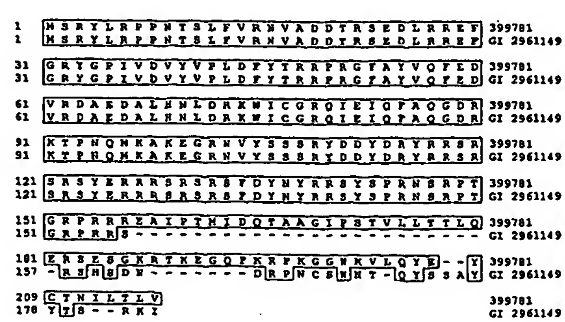
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: RNA-ASSOCIATED PROTEINS

(57) Abstract

The invention provides human RNA-associated proteins (RNAAP) and polynucleotides which identify and encode RNAAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonist. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of RNAAP.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

RNA-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of RNA-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, immune/inflammatory, and reproductive disorders.

BACKGROUND OF THE INVENTION

Ribonucleic acid (RNA) is a linear single-stranded polymer of four ribonucleotides, ATP, CTP, UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA), the genetic material of the organism. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function. Messenger RNAs (mRNAs) encode polypeptides. Ribosomal RNAs (rRNAs) are structural RNAs that are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that function in the translation of mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA translation by recognizing both an mRNA codon and the amino acid that matches that codon. Heterogeneous nuclear RNAs (hnRNAs) include mRNA precursors and other nuclear RNAs of various sizes. Small nuclear RNAs (snRNAs) are a part of the nuclear spliceosome complex that removes intervening, non-coding sequences (introns) and rejoins exons in pre-mRNAs.

RNA-binding proteins are essential for a wide variety of cellular and developmental functions. They participate in RNA processing, editing, transport, localization, stabilization, and the posttranscriptional control of mRNAs. They also provide the protein component of ribosomal RNA, transfer RNA, and other ribonuclear proteins. The RNA binding activity of these proteins is mediated by specific RNA-binding domains contained within the proteins. A variety of conserved RNA binding motifs have been defined through comparisons of amino acid homologies and structural similarities within these RNA-binding domains. These motifs include the RNP motif, an arginine-rich motif, the zinc-finger motif, the Y-box, the KH motif, and the double-stranded RNA-binding domain (dsRBD), all of which are characterized by specific consensus sequences (Burd, C. G. and Dreyfuss, G. (1994) Science 265:615 - 621).

<u>RNA Processing</u>

Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The spliceosomal complex is comprised of five small

nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base-pair with intron consensus sequences. The protein components mediate spliceosome assembly and the splicing reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, L. (1995) Biochemistry W.H. Freeman and Company, New York NY, p. 863).

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have roles in functions that include splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) Clin. Exp. Rheumatol. 16:317-326). Some examples of hnRNPs include the yeast proteins Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) Genes Dev. 12:679-691). A common feature of all of these RNA-binding proteins is a glycine-rich region in the form of RGG repeats. HnRNPs have been shown to be important targets of the autoimmune response in rheumatic diseases (Biamonti et al., supra).

An important means of regulating the function of hnRNPs is by methylation of arginine residues. The hnRNPs contain 65% of the methylated arginine residues in the cell nucleus. Methylation occurs within the RGG domain. Methylated arginine residues are also found in non-hnRNP RNA-binding proteins, all of which contain RGG repeats. The yeast enzyme, Hmt1p, is responsible for methylation of Npl3p and Hrp1p. In HMT1 null mutants, methylation of these proteins is not detectable, and poly(A^+)RNA accumulates in the nucleus. A molecular model predicts that Cbp80, Npl3p, and Hrp1p form a complex with mRNA to package the RNA for export from the nucleus, and that methylation plays a role in the efficiency of this packaging. Formation of this export complex is crucial for efficient exit of mRNA out of the nucleus. (Shen, supra.) A human homolog of Hmt1p, HRMT1L2, has been identified and is required for methylation of arginine residues in the RGG repeats of hnRNP A1. (Scott, H.S. et al. (1998) Genomics 48:330-340.) Also, viral RNA-binding proteins, such as the herpes simplex virus ICP27 protein, are known to be arginine-methylated. This exploitation of the cellular export machinery may facilitate maturation of viral RNAs. (Shen, supra.)

Human myxoid liposarcomas have been shown to contain a chromosomal translocation [(t12;16)(q13;p11)] wherein the gene coding for an inhibitory, growth arrest-associated transcription factor, known as CHOP (C/EBP homologous protein), is fused to the gene for TLS (translocated in liposarcoma), a nuclear RNA-binding protein that contains an RNP motif. TLS has been shown to function as an RNA chaperone, shuttling RNA into and out of the nucleus

(Zinszner, H. et al. (1997) J. Cell Sci. 110:1741-1450). The fusion of TLS with CHOP serves to convert a transcription factor involved in growth arrest into one associated with abnormal cell proliferation (Crozat, A. et al. (1993) Nature 363:640-644). Subsequent work has shown that TLS and its homologs (e.g., EWS, associated with Ewing's sarcoma) comprise the N-terminal portion of a number of fusion oncoproteins associated with sarcomas as well as with certain human acute myeloid leukemias (AMLs), secondary AMLs associated with myelodysplastic syndrome, and certain chronic myeloid leukemias (Aman, P. et al. (1996) Genomics 37:1-8; Zinszner, H. et al. (1997) Oncogene 14:451-461; Pereira, D.S. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8239-8244).

Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM) (Birney, E. et al. (1993) Nucleic Acids Res. 21:5803-5816). The RRM is about 80 amino acids in length and forms four β-strands and two α-helices arranged in an α/β sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to snRNP proteins, examples of RNA-binding proteins which contain the above motifs include heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative splicing. Alternative splicing factors include developmentally regulated proteins, specific examples of which have been identified in lower eukaryotes such as Drosophila melanogaster and Caenorhabditis elegans. These proteins play key roles in developmental processes such as pattern formation and sex determination, respectively (Hodgkin, J. et al. (1994) Development 120:3681-3689).

RNA Stability and Degradation

RNA helicases alter and regulate RNA conformation and secondary structure by using energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the "DEAD-box family," so named for the conserved B-type ATP-binding motif which is diagnostic of proteins in this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in various processes such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability. Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and embryogenesis. All DEAD-box helicases contain several conserved sequence motifs within about 420 amino acids. These motifs include an A-type ATP binding motif, the DEAD-box/B-type ATP-binding motif, a serine/arginine/threonine tripeptide of unknown function, and a C-terminal glycine-rich motif with a possible role in substrate binding and unwinding. In addition, alignment of divergent DEAD-box helicase sequences has shown that 37 amino acid residues are identical

among these sequences, suggesting that conservation of these residues is important for helicase function. (Reviewed in Linder, P. et al. (1989) Nature 337:121-122.) Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors, suggesting that DDX1 may promote or enhance tumor progression by altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in ultraviolet light-induced tumors, B-cell lymphoma, and myeloid malignancies (Godbout, R. et al. (1998) J. Biol. Chem. 273:21161-21168).

Ribonucleases (RNases) catalyze the hydrolysis of phosphodiester bonds in RNA chains, thus cleaving the RNA. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid, which occurs in cells invaded by retroviruses. RNase H is an important enzyme in the retroviral replication cycle. RNase H domains are often found associated with reverse transcriptases. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) Nat. Biotechnol. 15:529-536). Regulation of RNase activity may be a means for controlling tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

<u>Translation</u>

Proteins are translated from their RNA templates on the ribosome. The eukaryotic ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, the ribosome also contains more than fifty proteins. The ribosomal proteins have a prefix which denotes the subunit to which they belong, either L (large) or S (small). Three important sites are identified on the ribosome: i) the aminoacyl-tRNA site (A site) where charged tRNAs (except the initiator-tRNA) bind on arrival; ii) the peptidyl-tRNA site (P site) where new peptide bonds are formed and where the initiator tRNA binds, and iii) the exit site (E site) where deacylated tRNAs bind prior to their release from the ribosome (see Stryer, L. (1995) <u>Biochemistry</u> W.H. Freeman and Company, New York NY pp. 875-908; and Lodish, H. et al. (1995) <u>Molecular Cell Biology</u> Scientific American Books, New York NY pp. 119-138).

<u>tRNA Charging</u>

An important family of RNA-processing enzymes in the cytoplasm is the aminoacyl-transfer RNA (tRNA) synthetases. Protein biosynthesis depends on each amino acid forming a linkage with the appropriate tRNA. The aminoacyl-tRNA synthetases are responsible for correct attachment of an amino acid with its cognate tRNA. The 20 aminoacyl-tRNA synthetase enzymes

can be divided into two structural classes. each class characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding Rossman 'fold'. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet motif, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains. (Hartlein, M. and Cusack, S. (1995) J. Mol. Evol. 40:519-530.)

One of the best studied of the aminoacyl-tRNA synthetases is seryl-tRNA synthetase (SerRS). SerRS is a class II enzyme with an N-terminal regulatory domain in the form of a solvent exposed, antiparallel coiled-coil (the "helical arm"). A multiple sequence alignment and similarity plot of SerRS enzymes from prokaryotes, such as E. coli, and eukaryotes, such as yeast and mice, demonstrate the greatest variability in the N-terminal helical arm domain. Eukaryotic SerRS enzymes also contain a 20-48 amino acid C-terminal extension not found in prokaryotic synthetases. Truncation of the N-terminal helical arm causes SerRS to lose specificity for serine-tRNA. such that the truncated SerRS reacts with non-cognate tRNAs as well. In eukaryotes, loss of the C-terminal sequence does not have a major affect on enzymatic activity. (Hartlein, supra; and Weygand-Duraševic, I. et al. (1996) J. Biol. Chem. 271:2455-2461.)

Autoantibodies against aminoacyl-tRNAs are generated by patients with dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Translation Initiation

Initiation of translation can be divided into three stages. First an initiator transfer RNA (Met-tRNA$_f$) joins the 40S ribosomal subunit to form the 43S preinitiation complex. Next the 43S preinitiation complex binds the mRNA, and migrates to the correct AUG initiation codon. In the third step, the 60S ribosomal subunit joins the 40S subunit to generate an 80S ribosome at the inititation codon. Regulation of translation primarily involves the first and second stage in the initiation process (V.M. Pain (1996) Eur. J. Biochem. 236:747-771).

Several initiation factors, many of which contain multiple subunits, are involved in bringing an initiator tRNA and 40S ribosomal subunit together. eIF2B, a guanine nucleotide exchange protein, converts eIF2 from its GDP-bound inactive form to its GTP-bound active form. eIF2, a guanine nucleotide binding protein, recruits the initiator tRNA, bound to GTP, to the 40S ribosomal subunit. Two other factors, eIF1A and eIF3, bind and stabilize the 40S subunit by interacting with 18S ribosomal RNA and specific ribosomal structural proteins. eIF3 is also

involved in association of the 40S ribosomal subunit with mRNA. The Met-tRNA$_f$, eIF1A, eIF3, and 40S ribosomal subunit together make up the 43S preinitiation complex (Pain, supra).

Additional factors are required for binding of the 43S preinitiation complex to an mRNA molecule, and the process is regulated at several levels. eIF4F is a complex consisting of three proteins: eIF4E, eIF4A, and eIF4G. eIF4E recognizes and binds to the mRNA 5'-terminal m^{7}GTP cap, eIF4A is a bidirectional RNA-dependent helicase, and eIF4G is a scaffolding polypeptide. eIF4G has three binding domains. The N-terminal third of eIF4G interacts with eIF4E, the central third interacts with eIF4A, and the C-terminal third interacts with eIF3 bound to the 43S preinitiation complex. Thus, eIF4G acts as a bridge between the 40S ribosomal subunit and the mRNA (M.W. Hentze (1997) Science 275:500-501).

The ability of eIF4F to initiate binding of the 43S preinitiation complex is regulated by two structural features of the mRNA. The mRNA molecule has an untranslated region (UTR) between the 5' cap and the AUG start codon. In some mRNAs this region forms secondary structures that impede binding of the 43S preinitiation complex. Interestingly, the group of mRNAs possessing highly structured 5' UTRs includes a disproportionately high number of mRNAs encoding proteins that take part in or regulate processes involved in cell proliferation. The efficiency with which these mRNAs are translated may play a crucial role in the maintenance of correct restraints on cell growth. Additionally, regulatory proteins may bind to sites within the 5' UTR and stabilize this secondary structure to prevent translation. The helicase activity of eIF4A is thought to function in removing this secondary structure to facilitate binding of the 43S preinitiation complex (Pain, supra).

The second structural feature of mRNA regulating binding of the 43S preinitiation complex is the 3' poly(A) tail. The translational efficiency of an mRNA is related to the length of its poly(A) tail, such that the longer the tail the more efficient the translation of the message. This is due to an interaction between a protein that binds the poly(A) tail, the poly(A)-binding protein (PABP), and eIF4G. This interaction between PABP and eIF4G can only occur in the presence of RNA and involves a <120 amino acid site in the C-terminal half of eIF4G. This is an important form of regulation in translation of maternally-derived messages in early embryogenesis. The egg contains numerous mRNA molecules. Molecules with long poly(A) tails are translated early in development and then undergo poly(A) tail shortening to repress further translation. Messages with short poly(A) tails, which are initially left untranslated, go through a cytoplasmic tail elongation to initiate translation later in development. This process of tail length modification responds to developmental cues and also appears to involve PABP (Pain, supra).

Another level of regulation involving eIF4G has been demonstrated by infection of

mammalian cells with picornaviruses. Several members of the picornavirus family, including poliovirus, human rhinovirus 2, and foot-and-mouth disease virus, inhibit cellular mRNA translation by cleaving eIF4G into two fragments. This cleavage by the viral protease effectively separates the N-terminal eIF4E binding site from the C-terminal binding sites for eIF4A, eIF3, and PABP. Picornavirus RNAs, which are uncapped, utilize the C-terminal fragment of eIF4G for translation. This C-terminal fragment contains a region that interacts, either directly or indirectly, with an internal ribosome entry site (IRES) on the viral RNA molecule. Thus, eIF4G acts as a bridge between the 40S ribosome and the viral IRES for cap-independent translation as well (Hentze, supra).

Recently, a protein (p97) in yeast was shown to resemble the C-terminal fragment of eIF4G produced by picornavirus protease cleavage. p97 binds to both eIF3 and eIF4A, and may be involved in cap-independent translation of cellular mRNAs, though no candidate RNA species have been found within eukaryotic cells. p97 has been shown to be involved in modulating γ-interferon-induced programmed cell death (Hentze, supra).

Translation Elongation

Elongation, the joining of additional amino acids to the initiator methionine to complete the polypeptide chain, involves elongation factors EF1α, EF1βγ, and EF2. EF1α is a GTP-binding protein which, when bound by GTP, brings an aminoacyl-tRNA to the ribosome's A site. The amino acid attached to the newly arrived aminoacyl-tRNA forms a peptide bond with the initiatior methionine. The GTP on EF1α is hydrolyzed to GDP, and EF1α-GDP dissociates from the ribosome. EF1βγ binds EF1α -GDP and induces the dissociation of GDP from EF1α, allowing EF1α to bind GTP and a new cycle to begin.

As subsequent aminoacyl-tRNAs are brought to the ribosome, EF-G, another GTP-binding protein, catalyzes the translocation of tRNAs from the A site to the P site and finally to the E site of the ribosome.

Translation Termination

The release factor eRF carries out termination of translation. eRF recognizes stop codons in the mRNA, leading to the release of the polypeptide chain from the ribosome.

The discovery of new RNA-associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, immune/inflammatory, and reproductive disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, RNA-associated proteins,

referred to collectively as "RNAAP" and individually as "RNAAP-1," "RNAAP-2," "RNAAP-3," "RNAAP-4," "RNAAP-5," "RNAAP-6," "RNAAP-7," "RNAAP-8," "RNAAP-9," "RNAAP-10," "RNAAP-11," "RNAAP-12," "RNAAP-13," "RNAAP-14," "RNAAP-15," "RNAAP-16," and "RNAAP-17." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-17 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 and fragments thereof.

BRIEF DESCRIPTION OF FIGURES AND TABLES

Figure 1 shows the amino acid sequence alignment between RNAAP-1 (Incyte Clone number 399781; SEQ ID NO:1) and the human TLS-associated protein TASR (GI 2961149; SEQ ID NO:35), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 2A-H show the amino acid sequence alignment between RNAAP-2 (1252206; SEQ ID NO:2) and human eIF4G1 (GI 2660712; SEQ ID NO:36), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 3A and 3B show the hydropathy plots of RNAAP-2 (1252206; SEQ ID NO:2) and

human eIF4G1 (GI 2660712; SEQ ID NO:36), respectively. Plots were produced using MACDNASIS PRO software (Hitachi Software Engineering, S. San Francisco CA).

Figures 4A and 4B show the amino acid sequence alignment between RNAAP-3 (2950994; SEQ ID NO:3) and Drosophila seryl-tRNA synthetase (GI 2440051; SEQ ID NO:37), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Figures 5A-C show the amino acid sequence alignment between RNAAP-4 (3461657; SEQ ID NO:4) and human arginine methyltransferase (GI 1808648; SEQ ID NO:38), produced using the multisequence alignment program of LASERGENE software.

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding RNAAP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods and algorithms used for identification of RNAAP.

Table 3 shows useful fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding RNAAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze RNAAP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"RNAAP" refers to the amino acid sequences of substantially purified RNAAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which, when bound to RNAAP, increases or prolongs the duration of the effect of RNAAP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of RNAAP.

An "allelic variant" is an alternative form of the gene encoding RNAAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding RNAAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as RNAAP or a polypeptide with at least one functional characteristic of RNAAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding RNAAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding RNAAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent RNAAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity,

hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of RNAAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of RNAAP which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of RNAAP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to RNAAP, decreases the amount or the duration of the effect of the biological or immunological activity of RNAAP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of RNAAP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind RNAAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on

the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic RNAAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding RNAAP or fragments of RNAAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping

sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding RNAAP, by northern analysis is indicative of the presence of nucleic acids encoding RNAAP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding RNAAP.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR,

Madison WI) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) Parameters for each method may be the default parameters provided by MEGALIGN or may be specified by the user. The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which

may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of RNAAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of RNAAP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:18-34, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:18-34 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:18-34 from related polynucleotide sequences. A fragment of SEQ ID NO:18-34 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:18-34 and the region of SEQ ID NO:18-34 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" and "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding RNAAP, or fragments thereof, or RNAAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to

various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of RNAAP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to RNAAP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

THE INVENTION

The invention is based on the discovery of new human RNA-associated proteins (RNAAP), the polynucleotides encoding RNAAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, immune/inflammatory, and reproductive disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding RNAAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each RNAAP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The clones in column 5 were used to assemble the consensus nucleotide sequence of each RNAAP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows the identity of each polypeptide; and column 7 shows analytical methods used to identify each polypeptide through sequence homology and protein motifs. The segment of RNAAP-1 from residue R51 through residue D60, corresponding to region BL00030B, received a score of 1118 on a strength of 1104, while the segment from residue L12 through residue F30, corresponding to region BL00030A, received a score of 1089 on a strength of 1095, and supported the presence of BL00030B with a *P* value less than 2.4×10^{-4}.

As shown in Figure 1, RNAAP-1 has chemical and structural similarity with the human TLS-associated protein, TASR (GI 2961149; SEQ ID NO:35). In particular, RNAAP-1 and TASR share 76% identity, including the RNA recognition motif.

As shown in Figures 2 A-H, RNAAP-2 has chemical and structural similarity with human eIF4G1 (GI 2660712; SEQ ID NO:36). In particular, RNAAP-2 and human eIF4G1 share 45% identity and have similar isoelectric points (5.23 and 5.04, respectively). As shown in Figures 3A and 3B, RNAAP-2 and human eIF4G1 have similar hydrophobicity profiles.

As shown in Figures 4A and 4B, RNAAP-3 has chemical and structural similarity with Drosophila seryl-tRNA synthetase (GI 2440051; SEQ ID NO:37). In particular, RNAAP-3 and seryl-tRNA synthetase share 41% identity.

As shown in Figures 5A, 5B, and 5C, RNAAP-4 has chemical and structural similarity with human arginine methyltransferase (GI 1808648; SEQ ID NO:38). In particular, RNAAP-4 and arginine methyltransferase share 46% identity.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding RNAAP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1.

These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:18-34 and to distinguish between SEQ ID NO:18-34 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express RNAAP as a fraction of total tissues expressing RNAAP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing RNAAP as a fraction of total tissues expressing RNAAP. Northern analysis shows the expression of SEQ ID NO:18 in various libraries, at least 51% of which are associated with cancer and at least 29% of which are associated with inflammation and the immune response. Of particular note is SEQ ID NO: 29, which is expressed in only 25 libraries, 10(40%) of which are associated with reproductive tissue and 17(76%) of which are associated with cell proliferative disorders. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding RNAAP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The invention also encompasses RNAAP variants. A preferred RNAAP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the RNAAP amino acid sequence, and which contains at least one functional or structural characteristic of RNAAP.

The invention also encompasses polynucleotides which encode RNAAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34, which encodes RNAAP.

The invention also encompasses a variant of a polynucleotide sequence encoding RNAAP. In particular, such a variant polynucleotide sequence will have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding RNAAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34 which has at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:18-34. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of RNAAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the

genetic code, a multitude of polynucleotide sequences encoding RNAAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring RNAAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode RNAAP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring RNAAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding RNAAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding RNAAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode RNAAP and RNAAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding RNAAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:18-34 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and

most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 μg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art.

(See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding RNAAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal

using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode RNAAP may be cloned in recombinant DNA molecules that direct expression of RNAAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express RNAAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter RNAAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding RNAAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:225-232.) Alternatively, RNAAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of RNAAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active RNAAP, the nucleotide sequences encoding

RNAAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding RNAAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding RNAAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding RNAAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding RNAAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding RNAAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus,TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding RNAAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding RNAAP can

be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding RNAAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of RNAAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of RNAAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of RNAAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) Methods Enzymol. 153:516-54; and Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of RNAAP. Transcription of sequences encoding RNAAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding RNAAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses RNAAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of RNAAP in cell lines is preferred. For example, sequences encoding RNAAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ or *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding RNAAP is inserted within a marker gene sequence, transformed cells containing sequences encoding RNAAP can be identified by the absence of

marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding RNAAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding RNAAP and that express RNAAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of RNAAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on RNAAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul MN, Sect. IV; Coligan, J. E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding RNAAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding RNAAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding RNAAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The

protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode RNAAP may be designed to contain signal sequences which direct secretion of RNAAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas, VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding RNAAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric RNAAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of RNAAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the RNAAP encoding sequence and the heterologous protein sequence, so that RNAAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled RNAAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems

(Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of RNAAP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of RNAAP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of RNAAP and RNA-associated proteins. In addition, the expression of RNAAP is closely associated with reproductive tissues, nervous tissues, cell proliferation including cancer, and inflammation and immune response. Therefore, RNAAP appears to play a role in cell proliferative, immune/inflammatory, and reproductive disorders. In the treatment of disorders associated with increased RNAAP expression or activity, it is desirable to decrease the expression or activity of RNAAP. In the treatment of the above conditions associated with decreased RNAAP expression or activity, it is desirable to increase the expression or activity of RNAAP.

Therefore, in one embodiment, RNAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as
actinic keratosis, arteriosclerosis, atherosclerosis, bursitis. cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes

mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a reproductive disorder such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing RNAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified RNAAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of RNAAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of RNAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of RNAAP. Examples of such disorders include, but are not limited to, those described above. In one aspect, an antibody which specifically binds RNAAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express RNAAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding RNAAP may be administered to a subject to treat or prevent a disorder associated with

increased expression or activity of RNAAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of RNAAP may be produced using methods which are generally known in the art. In particular, purified RNAAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind RNAAP. Antibodies to RNAAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with RNAAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to RNAAP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of RNAAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to RNAAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-

hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce RNAAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for RNAAP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between RNAAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering RNAAP epitopes is preferred, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for RNAAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of RNAAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are

heterogeneous in their affinities for multiple RNAAP epitopes, represents the average affinity, or avidity, of the antibodies for RNAAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular RNAAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the RNAAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of RNAAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of RNAAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding RNAAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding RNAAP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding RNAAP. Thus, complementary molecules or fragments may be used to modulate RNAAP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding RNAAP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding RNAAP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding RNAAP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding RNAAP.

Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding RNAAP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding RNAAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding RNAAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or

SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of RNAAP, antibodies to RNAAP, and mimetics, agonists, antagonists, or inhibitors of RNAAP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which

facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include

fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of RNAAP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example RNAAP or fragments thereof, antibodies of RNAAP, and agonists, antagonists or inhibitors of RNAAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic

effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind RNAAP may be used for the diagnosis of disorders characterized by expression of RNAAP, or in assays to monitor patients being treated with RNAAP or agonists, antagonists, or inhibitors of RNAAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for RNAAP include methods which utilize the antibody and a label to detect RNAAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring RNAAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of RNAAP expression. Normal or standard values for RNAAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to

RNAAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of RNAAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding RNAAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of RNAAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of RNAAP, and to monitor regulation of RNAAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding RNAAP or closely related molecules may be used to identify nucleic acid sequences which encode RNAAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding RNAAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the RNAAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:18-34 or from genomic sequences including promoters, enhancers, and introns of the RNAAP gene.

Means for producing specific hybridization probes for DNAs encoding RNAAP include the cloning of polynucleotide sequences encoding RNAAP or RNAAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding RNAAP may be used for the diagnosis of disorders associated with expression of RNAAP. Examples of such disorders include, but are not limited to,

a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a reproductive disorder such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding RNAAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered RNAAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding RNAAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The

nucleotide sequences encoding RNAAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding RNAAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of RNAAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding RNAAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding RNAAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding RNAAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding RNAAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less

stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of RNAAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding RNAAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding RNAAP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene

sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, RNAAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between RNAAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with RNAAP, or fragments thereof, and washed. Bound RNAAP is then detected by methods well known in the art. Purified RNAAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding RNAAP specifically compete with a test compound for binding RNAAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with RNAAP.

In additional embodiments, the nucleotide sequences which encode RNAAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0598 P, filed September 22, 1998], U.S. Ser. No. [Attorney Docket No. PF-0600 P, filed September 17, 1998], U.S. Ser. No. [Attorney Docket No. PF-0626 P, filed November 4, 1998], and U.S. Ser. No. 60/128,660, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the

polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), pSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies. **II.**

Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those

skilled in the art. Table 5 summarizes the tools. programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used. the second column provides brief descriptions thereof, the third column presents appropriate references. all of which are incorporated by reference herein in their entirety, and the fourth column presents. where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Str. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:18-34. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7;

Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding RNAAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Extension of RNAAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:18-34 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art.

PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer

sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:18-34 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:18-34 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba1, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected

using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the RNAAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring RNAAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of RNAAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the RNAAP-encoding transcript.

IX. Expression of RNAAP

Expression and purification of RNAAP is achieved using bacterial or virus-based expression systems. For expression of RNAAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express RNAAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of RNAAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding RNAAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional

genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, RNAAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from RNAAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified RNAAP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of RNAAP Activity

RNAAP activity is demonstrated by a polyacrylamide gel mobility-shift assay. In preparation for this assay, RNAAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing RNAAP cDNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of RNAAP. Extracts containing solubilized proteins can be prepared from cells expressing RNAAP by methods well known in the art. Portions of the extract containing RNAAP are added to [^{32}P]-labeled RNA. Radioactive RNA can be synthesized in vitro by techniques well known in the art. The mixtures are incubated at 25°C in the presence of RNase inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates the formation of a complex between RNAAP and the radioactive transcript. A band of similar mobility will be absent in samples prepared using control extracts prepared from untransformed cells.

Alternatively, the activity of RNAAP is measured as the level of in vitro translation of cap-dependent chloramphenicol acetyltransferase (CAT) and cap-independent luciferase (LUC) reporter constructs (Haghighat, A., et al. (1996) J. Virol. 70:8444-8450). Bicistronic pGEMCAT/EMC/LUC mRNA is used in the assay. The first cistron on this mRNA construct encodes the CAT protein and its translation is cap-dependent. The second cistron encodes luciferase enzyme. The encoded region of the second cistron is preceded by the IRES of

encephalomyocarditis (EMC) virus, making luciferase translation cap independent. Linearized pGEMCAT/EMC/LUC is transcribed in vitro using T7 RNA polymerase in the presence of 10-fold molar excess m^7GpppG, a cap analog that promotes capping of the RNA product. Rabbit reticulocyte lysate is treated with picornavirus 2A protease. Treatment of the lysate with 2A protease reduces cap-dependent (CAT) translation, but does not inhibit cap-independent (luciferase) translation. Treated lysate is programmed by addition of the capped mRNA in the presence of 20 μCi [^{35}S]methionine. Translation reaction mixtures are incubated for 90 min in the presence of added eIF4E, RNAAP, eIF4E and RNAAP, or with no additions. Translation products are analyzed by SDS-PAGE, acid fixation, and autoradiography. RNAAP activity is calculated based on the expression level of CAT relative to luciferase as compared to control reactions lacking RNAAP.

Alternatively, RNAAP activity is measured as the aminoacylation of a substrate tRNA in the presence of [^{14}C]serine. RNAAP is incubated with $tRNA^{ser}$ and [^{14}C]serine in a buffered solution. ^{14}C-labeled product is separated from free [^{14}C]serine by chromatography, and the incorporated ^{14}C is quantified by scintillation counter. The amount of ^{14}C detected is proportional to the activity of RNAAP in this assay.

Alternatively, RNAAP activity is measured as the methylation of a substrate in the presence of [methyl-^{3}H]-*S*-adenosylmethionine (SAM). RNAAP is incubated with an appropriate substrate and [methyl-^{3}H]SAM in a buffered solution. ^{3}H-labeled product is separated from free [methyl-^{3}H]SAM by gel electrophoresis, and the incorporated ^{3}H is quantified by fluorography. The amount of ^{3}H detected is proportional to the activity of RNAAP in this assay.

XI. Functional Assays

RNAAP function is assessed by expressing the sequences encoding RNAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion

protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of RNAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding RNAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding RNAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of RNAAP Specific Antibodies

RNAAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the RNAAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit

antisera. washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring RNAAP Using Specific Antibodies

Naturally occurring or recombinant RNAAP is substantially purified by immunoaffinity chromatography using antibodies specific for RNAAP. An immunoaffinity column is constructed by covalently coupling anti-RNAAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing RNAAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of RNAAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/RNAAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and RNAAP is collected.

XIV. Identification of Molecules Which Interact with RNAAP

RNAAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled RNAAP, washed, and any wells with labeled RNAAP complex are assayed. Data obtained using different concentrations of RNAAP are used to calculate values for the number, affinity, and association of RNAAP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	18	399781	PITUNOT02	399781H1 and 399781X12 (PITUNOT02), 1271965F6 (TESTTUT02), 790764R1 and 792124R1 (PROSTUT03), and 405935R1 (EOSIHET02)
2	19	1252206	LUNGFET03	1232931T6 (LUNGFET03), 3109423H1 (BRSTTUT15), 3113355H1 (BRSTNOT17), 3330287H1 (HEAONOT04), 3269650H1 (BRAINOT20), 1662596H1 (BRSTNOT09), 2655078H1 (THYMNOT04), 2266829H1 and 2266829R6 (UTRSNOT02), 4333545H1 (KIDCTMT01), 1595462F6 (BRAINOT14), 078192R1 and 078192F1 (SYNORAB01), 4836680H1 (BRAWNOT01), 1252206F6 (LUNGFET03), 1638473F6 (UTRSNOT06), SAJA00661R1, SAJA00355F1, SAJA01106R1, SAJA01874F1, and SAJA02468F1
3	20	2950994	KIDNFET01	1968448H1 (BRSTNOT04), 1435425T6 (PANCNOT08), 808869T1 (LUNGNOT04), 2795721F6 (NPOLNOT01), and 2950994H1 (KIDNFET01)
4	21	3461657	293TF201	2606248F6 (LUNGTUT07), 2052041X301D1 (LIVRFET02), 4341820F6 (BRAUNOT02), 2789769F6 (COLNTUT16), 3461657H1 (293TF2T01), SBUA03574D1 and SBUA00296D1
5	22	053076	FIBRNOT01	053076H1 (FIBRNOT01), 534171F1 (BRAINOT03), 4717220H1 (BRAIHCT02)
6	23	1292379	PGANNOT03	458715T6 (KERANOT01), 850050T1 (NGANNOT01), 1292379F1, 1292379H1 and 1292379T1 (PGANNOT03), 1398840F6 and 1398840T6 (BRAITUT08), 3447383H2 (BLADNOT09), 3780263H1 (BRSTNOT27)
7	24	1437783	PANCNOT08	117781F1 (KIDNNOT01), 1352071F1 (LATRTUT02), 1437783H1 (PANCNOT08), 2527706H1 (GBLANOT02), 4567705H1 (HELATXT01)
8	25	1557635	BLADTUT04	077627R1 (SYNORAB01), 1557635F1 and 1557635H1 (BLADTUT04), 1568446F1 (UTRSNOT05), 1901128F6 (BLADTUT06), 2013353T6 (TESTNOT03), 2098109H1 (BRAITUT02), 2568583T6 (HIPOAZT01), 3866538H1 (BRAITUT07)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
9	26	2049352	LIVRFET02	078075R1 (SYNORAB01), 994247R6 (COLNNOT11), 1334674F6 (COLNNOT13), 2049352F6 and 2049352H1 (LIVRFET02), 3219182H1 (COLNNON03)
10	27	2231663	PROSNOT16	307827H1 (HEARNOT01), 1455948F1 and 1455948R1 (COLNFET02), 2231663H1 (PROSNOT16), 3779128H1 (BRSTNOT27)
11	28	2604449	LUNGTUT07	606296R6 (BRSTTUT01), 1718568T6 (BLADNOT06), 2604449F6 and 2604449H1 (LUNGTUT07), 5093027F6 (UTRSTMR01), SAEA01050F1, SAEA01365F1, SAEC11108F1, SBKA00681F1
12	29	2604993	LUNGTUT07	1441072F6 and 1441072T6 (THYRNOT03), 2604993H1 (LUNGTUT07), 3389190T6 (LUNGTUT17), SBIA05937D1, SBIA11687D1, SBIA04881D1, SBIA03937D1, SBIA00985D1
13	30	2879070	UTRSTUT05	1458387F7, 1458387R1, and 1458387T6 (COLNFET02), 1858014X13C1 and 1858014X14C1 (PROSNOT18), 2595610H1 (OVARTUT02), 2879070H1 (UTRSTUT05)
14	31	3093845	BRSTNOT19	134421R1 (BMARNOT02), 979683R6 (TONGTUT01), 3093845F6 and 3093845H1 (BRSTNOT19), 3294785F6 (TLYJINT01)
15	32	3685685	HEAANOT01	1556450F1 (BLADTUT04), 1615712T6 (BRAITUT12), 2041291R6 (HIPONON02), 2448460F6 (THP1NOT03), 3685685H1 (HEAANOT01), 3954790H1 (PONSAZT01), 4918977H2 (TESTNOT11)
16	33	3825977	BRAINOT23	2373839T6 and 2375912X302D1 (ISLTNOT01), 3825977H1 (BRAINOT23), 3882790H1 (SPLNNOT11), SBIA02579D1, SBIA02994D1, SBIA10082D1, SBIA06183D1, SBIA05526D1, SBIA02807D1
17	34	4941262	BRAIFEN03	4941262F6 and 4941262H1 (BRAIFEN03)

Table 2

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
1	216	S129, T21, S108, T161, T178, T47, S107, S143, T150, S185,Y116, Y138	N9	RNA recognition motif: L12-I83 RNA-binding region RNP-1 R51 signature: R1-D60, L12-F30	GI 2961149 Hhuman TLS-associated protein, TASR	Motifs BLAST PFAM BLOCKS
2	1584	S740, S888, S965, S257, T294, S304, S317, S366, S370, T517, S542, S582, S584, S598, T615, S718, S865, T1058, S1085, T1115, S1155, S1164, T1190, S1209, S1217, S1227, S1264, S1290, S1333, S1381, S1416, S1421, S1501, T1503, S1550, S30, T141, S304, S362, S456, S491, T507, S611, S700, S718, S735, T817, S965, S985, S1121, T1126, T1144, S1155, T1175, S1200, S1286, S1333, S1367, S1381, S1416, T1480, S1550	N1162, N1188, N1195	Leucine zipper pattern: L1513-L1534 Wilm's tumor protein: G80-P94, S412-H426	GI 2660712 Human eIF4G1	Motifs BLAST PRINTS

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
3	166	S78, T135	N72, N99		GI 2440051 seryl-tRNA synthetase	BLAST
4	531	S27, T58, S59, S157, S242, S339, S428, S430, S242, T439, S475, S492, Y89	N155, N522, N523	C2H2 type zinc finger motif: C50-H71 N-methyltransferase cofactor-binding motif: V259-A273	GI 1808648 Human arginine methyltransferase	Motifs BLAST BLOCKS PFAM PRINTS
5	148	S32 S38 S47 T69 T141 Y60		A31-D115 (Ribosomal L27 protein) M1-A27 (Signal peptide)	ribosomal protein L27 g 642605	Motifs BLAST Pfam HMM SPScan
6	317	S20 S40 S106 S110 S117 T135 T142 S144 T260 S302 S6 S10 T134 S215 S281	N148 N208 N228		pre-ribosomal particle assembly protein g 2398808	Motifs BLAST
7	278	T10 S83 S56 T57 T61 T121 S202 S244 T13 T68 T156 T192 S224 Y251	N71 N120		translation initiation factor 3 (infC) g 3844793	Motifs BLAST

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
8	586	T29 T81 T261 S512 T4 S21 T29 S97 S227 T229 S235 T348 S371 S417 T475 T485 S511 S513 S515 S554 T562 S77 T127 T194 S206 S215 S256 S356 S479 Y274 Y297 Y309	N427		Similar to mRNA splicing factor g 3878326	Motifs BLAST
9	384	T32 S167 T327 T339 T349 S28 T148 T311 S372 Y13 Y19 Y86 Y277	N229	H257-M296 (Cytidine and deoxycytidylate deaminases zinc-binding region signature)	phorbolin I protein kinase C associated protein g 436941	Motifs BLAST
10	325	T61 S298 S320 S49 T53 S116	N163	R94-G302 (L1P family ribosomal proteins)	Ribonucleotide reductase subunit M2 g 200768	Motifs BLAST Pfam

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential glycosylation sites	Signature Sequence	Identification	Analytical Methods
11	351	S39 T182 S329 S18 S29 T65 T182 S225 S38 Y87	N23 N314	E131-I146 (Ribonucleotide reductase small subunit) P46-D100, F123-D148, F198-F239, V251-R292 (Ribonucleotide reductase) W69-Y331 (Ribonucleotide reductase) R186-W207 (transmembrane)	Ribonucleotide reductase subunit M2 g 200468	Motifs BLAST Pfam BLOCKS HMM
12	681	T68 S79 S135 T160 S179 S201 S216 S237 T301 T312 T338 T363 T405 T457 S524 S123	N89 N600 N623	V227-V297, V328-L401, I447-V520 (RNA recognition motif) M1-K22 (signal peptide)	Similarity to Human heterogeneous nuclear ribonucleoprotein(hnRNP) F protein g 3880146	Motifs BLAST Pfam SPScan
13	408	S3 S45 S68 T212 T236 S248 T145 T279 Y193	N206	I121-M144 (transmembrane)	RNA helicase A g2880057	Motifs BLAST HMM

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential glycosylation sites	Signature Sequence	Identification	Analytical Methods
14	351	S126 S5 T7 S75 S108 S140 S195 S314 S339 S59 S122 S254 S300 S344 Y23	N113 N202	K36-Y43 (Eukaryotic putative RNA-binding region RNP-1 signature) I2-L38, V127-V194, L269-V334 (RNA recognition motif)	Hel-N2 RNA binding protein g905387	Motifs BLAST Pfam
15	472	S69 S116 S346 S89 S237 S239 S301 T303 S358 S4 T39 S124 T176	N219 N248	102-130, 178-204 (glycosyl hyrolase)	Human RNA binding protein g 2804465	Motifs BLAST PRINTS
16	616	S154 S368 S376 T570 S14 S44 T53 S83 S94 S466		V18-V89 (RNA recognition motif) F36-R85 (eukaryotic RNA-binding RNP-1)	Cleavage stimulating factor g 181139	Motifs BLAST Pfam ProfileScan
17	112	T42 Y69		G74-P95 (ribosomal protein L35Ae signature) L12-F106 (ribosomal protein L35Ae signature)	g4392 ribosomal protein L37a	Motifs BLAST Pfam BLOCKS

Table 3

Polynucleotide SEQ ID NO:	Selected Fragment (Nucleotide number)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
18	30-90	Nervous (0.191) Reproductive (0.309)	Cell proliferation (0.510) Inflammation and Immune Response (0.290)	PSPORT1
19	1137-1196	Nervous (0.245) Reproductive (0.216)	Cell proliferation (0.560) Inflammation and Immune Response (0.230)	pINCY
20	454-510	Reproductive (0.263) Nervous (0.211)	Cancer (0.580) Inflammation and Immune Response (0.160)	pINCY
21	31-81	Nervous (0.357) Gastrointestinal (0.179) Reproductive (0.143)	Cancer (0.610) Inflammation and Immune Response (0.210)	pINCY
22	1-46	Reproductive (0.247) Nervous (0.183) Gastrointestinal (0.118)	Cell proliferation (0.613) Inflammation (0.290)	PBLUESCRIPT
23	273-317	Reproductive (0.256) Nervous (0.209)	Cell proliferation (0.465) Inflammation (0.256)	pINCY
24	434-478	Gastrointestinal (0.244) Nervous (0.186) Reproductive (0.163)	Cell proliferation (0.535) Inflammation (0.361)	pINCY
25	174-218	Reproductive (0.230) Nervous (0.216) Cardiovascular (0.122)	Cell proliferation (0.554) Inflammation (0.311)	pINCY

Table 3 (cont.)

Polynucleotide SEQ ID NO:	Selected Fragment (Nucleotide number)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
26	489-533	Reproductive (0.270) Hematopoietic/Immune (0.243) Nervous (0.162)	Cell proliferation (0.676) Inflammation (0.405)	pINCY
27	199-252	Reproductive (0.308) Cardiovascular (0.205)	Cell proliferation (0.770) Inflammation (0.128)	pINCY
28	110-154	Cardiovascular (0.289) Nervous (0.184) Reproductive (0.158)	Cell proliferation (0.685) Inflammation (0.158)	pINCY
29	326-370	Reproductive (0.400) Gastrointestinal (0.240) Cardiovascular (0.120)	Cell proliferation (0.760) Inflammation (0.240)	pINCY
30	516-563	Reproductive (0.415) Nervous (0.151) Hematopoietic/Immune (0.113)	Cell proliferation (0.566) Inflammation (0.320)	pINCY
31	272-316	Hematopoietic/Immune (0.286) Gastrointestinal (0.214) Reproductive (0.214)	Inflammation (0.714) Cell proliferation (0.495)	pINCY
32	119-163	Reproductive (0.328) Hematopoietic/Immune (0.219) Nervous (0.156)	Cell proliferation (0.672) Inflammation (0.313)	pINCY
33	812-856	Gastrointestinal (0.208) Hematopoietic/Immune (0.208) Developmental (0.167) Nervous (0.167)	Inflammation (0.541) Cell proliferation (0.458)	pINCY
34	42-86	Nervous (1.000)	Cell proliferation (1.000)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
18	PITUNOT02	Library was constructed using RNA isolated from the pituitary glands removed from a pool of 87 male and female donors, 15 to 75 years old (RNA acquired from Clontech, CLON 6584-1).
19	LUNGFET03	Library was constructed RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation. Family history included bronchitis.
20	KIDNFET01	Library was constructed using RNA isolated from kidney tissue removed from a Caucasian female fetus, who died at 17 weeks' gestation from anencephalus.
21	293TF201	Library was constructed using RNA isolated from a treated, transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were treated with 5-aza-2'-deoxycytidine (5AZA) and transformed with adenovirus 5 DNA.
22	FIBRNOT01	Library was constructed using RNA isolated from the WI38 lung fibroblast cell line, which was derived from a 3-month-old Caucasian female fetus. $2x10^{-6}$ primary clones were then amplified to stabilize the library for long-term storage.
23	PGANNOT03	Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule.
24	PANCNOT08	Library was constructed using RNA isolated from pancreatic tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology for the associated tumor tissue indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Family history included cardiovascular disease, type II diabetes, and stomach cancer.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
25	BLADTUT04	Library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Patient history included tobacco use. Family history included type I diabetes, malignant neoplasm of the stomach, atherosclerotic coronary artery disease, and an acute myocardial infarction.
26	LIVRFET02	Library was constructed using RNA isolated from liver tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation. Family history included bronchitis.
27	PROSNOT16	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA) and was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
28	LUNGTUT07	Library was constructed using RNA isolated from lung tumor tissue removed from the upper lobe of a 50-year-old Caucasian male during segmental lung resection. Pathology indicated an invasive grade 4 squamous cell adenocarcinoma. Patient history included tobacco use. Family history included skin cancer.
29	LUNGTUT07	Library was constructed using RNA isolated from lung tumor tissue removed from the upper lobe of a 50-year-old Caucasian male during segmental lung resection. Pathology indicated an invasive grade 4 squamous cell adenocarcinoma. Patient history included tobacco use. Family history included skin cancer.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
30	UTRSTUT05	Library was constructed using RNA isolated from uterine tumor tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated uterine leiomyoma. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Patient history included a ventral hernia and a benign ovarian neoplasm.
31	BRSTNOT19	Library was constructed using RNA isolated from breast tissue removed from a 67-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated residual invasive lobular carcinoma. The focus of residual invasive carcinoma was positive for both estrogen and progesterone. Patient history included depressive disorder and benign large bowel neoplasm. Family history included cerebrovascular disease, benign hypertension, congestive heart failure, and lung cancer.
32	HEAANOT01	Library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco use. Family history included atherosclerotic coronary artery disease.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
33	BRAINOT23	Library was constructed using RNA isolated from right temporal lobe tissue removed from a 45-year-old Black male during a brain lobectomy. Pathology for the associated tumor tissue indicated dysembryoplastic neuroepithelial tumor of the right temporal lobe. The right temporal region dura was consistent withc alcifying pseudotumor of the neuraxis. The patient presented with convulsive intractable epilepsy, partial epilepsy, and memory disturbance. Patient history included obesity, meningitis, backache, unspecified sleep apnea, acute stress reaction, acquired knee deformity, and chronic sinusitis. Family history included obesity, benign hypertension, cirrhosis of the liver, alcohol abuse, hyperlipidemia, cerebrovascular disease, and type II diabetes.
34	BRAIFEN03	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus with a hypoplastic left heart stillborn after 23 weeks' gestation. The library was normalized in two rounds (with 48 hour reannealing hybridizations) using conditions adapted from Soares et al. and Bonaldo et al.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences:* Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score= 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ

ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.

12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

18. A purified antagonist of the polypeptide of claim 1.

19. A method for treating or preventing a disorder associated with decreased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a disorder associated with increased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

```
1    MSRYLRPPNTSLFVRNVADDTRSEDLRREF   399781
1    MSRYLRPPNTSLFVRNVADDTRSEDLRREF   GI 2961149

31   GRYGPIVDVYVPLDFYTRRPRGFAYVQFED   399781
31   GRYGPIVDVYVPLDFYTRRPRGFAYVQFED   GI 2961149

61   VRDAEDALHNLDRKWICGRQIEIQFAQGDR   399781
61   VRDAEDALHNLDRKWICGRQIEIQFAQGDR   GI 2961149

91   KTPNQMKAKEGRNVYSSSRYDDYDRYRRSR   399781
91   KTPNQMKAKEGRNVYSSSRYDDYDRYRRSR   GI 2961149

121  SRSYERRRSRSRSFDYNYRRSYSPRNSRPT   399781
121  SRSYERRRSRSRSFDYNYRRSYSPRNSRPT   GI 2961149

151  GRPRRREAIPTMIDQTAAGIPSTVLLTTLQ   399781
151  GRPRRS------------------------   GI 2961149

181  ERSESGKRTKEGQFKRPKGGWKVLQYE--Y   399781
157  -RSHSDN--------DRPNCSWNT-QYSSAY   GI 2961149

209  CTNILTLV                         399781
178  YTS--RKI                         GI 2961149
```

FIGURE 1

```
1    M N S Q P Q T R S P F F Q R P Q I Q P P R A T I P N S S P S  1252206
1    M S G A R T A S T P - - - - - - - - - - - - T P P Q T G G G  GI 2660712

31   I R P G A Q T P T A V Y Q A N Q H I M M V N H L P M P Y P V  1252206
19   L E P Q A N G E T P - - - - - - Q V A V I V R - - - - - - -  GI 2660712

61   P Q G P Q Y C I P Q Y R H S G P P Y V G P P Q K Y P V Q P P  1252206
36   - - - - - - - - P D D R S Q G A I I A D R P G - - - - - - -  GI 2660712

91   G P G P F Y P G P G P G D F P N A Y G T P F Y P S Q P V Y Q  1252206
51   - - - - - L P G P E H S - - - - - - - - - - - - - - - - - -  GI 2660712

121  S A P I I V P T Q Q Q P P P A K R E K K T I R I R D P N Q G  1252206
58   - - - - - - P S E S Q P S S P S P T P S P S P V L E P - - -  GI 2660712

151  G K D I T E E I M S G G G S R N P T P P I G R P T S T P T P  1252206
79   - - - - - - - - - - - - G S E P N L A V L S I P G D T M T T  GI 2660712

181  P Q L P S Q V P E H S P V V Y G T V E S A H L A A S T P V T  1252206
97   I Q M S - - V E E S T P I S R E T G E P Y R L S - - - - - -  GI 2660712
```

FIGURE 2A

```
211  AASDQKQEEKPKPDPVLKSPSPVLRLVLSG  1252206
119  ----------PEPTPL---AEPILEVEVTL  GI 2660712

241  EKKEQEGQTSETTAIVSIAELPLPPSPTTV  1252206
136  SKPVPESEFSSS---------PLQAP----  GI 2660712

271  SSVARSTIAAPTSSALSSQPIFTTAIDDRC  1252206
153  TPLASHTVEIHEPNGMVPSEDLEPEVESSP  GI 2660712

301  ELSSPREDTIPIPSLTSCTETSDPLPTNEN  1252206
183  ELAPPP----------ACPSES--------  GI 2660712

331  DDDICKKPCSVAPNDIPLVSSTNLINEING  1252206
195  -------PVPIAPTAQP-------------  GI 2660712

361  VSEKLSATESIVEIVKQEVLPLTLELEILE  1252206
205  -----------EELLNGAPSPPAVDLSPVS  GI 2660712

391  NPPEEMKLECIPAPITPSTVPSFPPTPPTP  1252206
224  EPEEQAKE--VTASVAPPTIPSATPATAPS  GI 2660712
```

FIGURE 2B

```
421  P A S P P H T P V I V P A A A T T V S S P S A A I T V Q R V  1252206
252  A T S P A Q E E E M E E E E E E E E G E A G E A G E A E S E  GI 2660712

451  L E E D E S I R T C L S E D A K E I Q N K I E V E A D G Q T  1252206
282  K G G E E L - - - - L P P E S T P I P A N L - - - - - - - -  GI 2660712

481  E E I L D S Q N L N S R R S P V P A Q I A I T V P K T W K K  1252206
300  - - - - - S Q N L E A - - - A A A T Q V A V S V P K R R R K  GI 2660712

511  P K D R T R T T E E M L E A E L E L K A E E E L S I D K V L  1252206
322  I K E L N K K - - E A V G D L L D A F K E A N P A V P E V -  GI 2660712

541  E S E Q D K M S Q G F H P E R D P S D L K K V K A V E E N G  1252206
349  - - - E N Q P P A G S N P G P E S E G - - - - S G V P P R P  GI 2660712

571  E E A E P V R N G A E S - V S E G E G I D A N S G S T D S S  1252206
372  E E A D E T W D S K E D K I H N A E N I Q P G E Q K - - - -  GI 2660712

600  G D G V T F P F K P E S W K P T D T E G K K Q Y D R E F L L  1252206
398  - - - - - Y E Y K S D Q W K P P N L E E K K R Y D R E F L L  GI 2660712
```

FIGURE 2C

```
630  D F Q F M P A C I Q K P E G L P P I S D V V L D K I N Q P K  1252206
423  G F Q F I F A S M Q K P E G L P H I S D V V L D K A N - - K  GI 2660712

660  L P M R T L D P R I L P R - - - G P D F T P A F A D F G R Q  1252206
451  T P L R P L D P T R L Q G I N C G P D F T P S F A N L G R T  GI 2660712

687  T P G G R G V P - - - - - - - - - - - - L L N V G S R R S Q  1252206
481  T L S T R G P P R G G P G G E L P R G P Q A G L G P R R S Q  GI 2660712

705  P G Q R R E P R K I I - T V S V K E D V H L K K A E N A W K  1252206
511  Q G P R K E P R K I I A T V L M T E D I K L N K A E K A W K  GI 2660712

734  P S Q K R - - - - - D S Q A D D P E N I K T Q E L F R K V R  1252206
541  P S S K R T A A D K D R G E E D A D G S K T Q D L F R R V R  GI 2660712

759  S I L N K L T P Q M F N Q L M K Q V S G L T V D T E E R L K  1252206
571  S I L N K L T P Q M F Q Q L M K Q V T Q L A I D T E E R L K  GI 2660712

789  G V I D L V F E K A I D E P S F S V A Y A N M C R C L V T L  1252206
601  G V I D L I F E K A I S E P N F S V A Y A N M C R C L M A L  GI 2660712
```

FIGURE 2D

```
819  K V P M A D K P G N T V N F R K L L L N R C Q K E F E K D K  1252206
631  K V P T T E K P T V T V N F R K L L L N R C Q K E F E K D K  GI 2660712

849  A D D D V F E K K Q K E L E A A S A P E E R T R L H D E L E  1252206
661  D D D E V F E K K Q K E M D E A A T A E E R G R L K E E L E  GI 2660712

879  E A K D K A R R R S I G N I K F I G E L F K L K M L T E A I  1252206
691  E A R D I A R R R S L G N I K F I G E L F K L K M L T E A I  GI 2660712

909  M H D C V V K L L K N H D E E S L E C L C R L L T T I G K D  1252206
721  M H D C V V K L L K N H D E E S L E C L C R L L T T I G K D  GI 2660712

939  L D F E K A K P R M D Q Y F N Q M E K I V K E K K T S S R I  1252206
751  L D F E K A K P R M D Q Y F N Q M E K I I K E K K T S S R I  GI 2660712

969  R F M L Q D V I D L R L C N W V S R R A D Q G P K T I E Q I  1252206
781  R F M L Q D V L D L R G S N W V P R R G D Q G P K T I D Q I  GI 2660712

999  H K E A K I E E Q E E Q R K V Q Q L M T K - - E K R R - - -  1252206
811  H K E A E M E E H R E H I K V Q Q L M A K G S D K R R G G P  GI 2660712
```

FIGURE 2E

```
1024  --PGVQR----VDEGGWNTVQGAKNSRVLD  1252206
841   PGPPISRGLPLVDDGGWNTVPISKGSRPID  GI 2660712

1048  PSKFLKITKP-TIDEKIQLVPKAQLGSWGK  1252206
871   TSRLTKITKPGSIDSNNQLFAPGGRLSWGK  GI 2660712

1077  GSSGG--AKASET--DALRSSASSLNRFSA  1252206
901   GSSGGSGAKPSDAASEAARPATSTLNRFSA  GI 2660712

1103  LQPPAPSGSTPSTPVEFDSRRTLTSRGSMG  1252206
931   LQQAVPTEST--------DNRRVVQ-RSSLS  GI 2660712

1133  REKNDKPLPSATARPNTFMRGGSSKDLLDN  1252206
953   RERGEKA-GDRGDRLERSERGGDRGDRLDR  GI 2660712

1163  QSQEEQRR----EMLETVKQLTGGVDVERN  1252206
982   ARTPATKRSFSKEVEERSRERPSQPEGLRK  GI 2660712

1189  STE--AERNKTRESAKPEISAMSAHD-KAA  1252206
1012  AASLTEDRDRGRDAVKREAALPPVSPLKAA  GI 2660712
```

FIGURE 2F

```
1216 LSEEELERKSKSIIDEFLHINDFKEAMQCV 1252206
1042 LSEEELEKKSKAIIEEYLHLNDMKEAVQCV GI 2660712

1246 EELNAQGLLHVFVRVGVESTLERSQITRDH 1252206
1072 QELASPSLLFIFVRHGVESTLERSAIAREH GI 2660712

1276 MGQLLYQLVQSEKLSKQDFFKGFSETLELA 1252206
1102 MGQLLHQLLCAGHLSTAQYYQGLYEILELA GI 2660712

1306 DDMAIDIPHIWLYLAELVTPMLKEGGISMR 1252206
1132 EDMEIDIPHVWLYLAELVTPILQEGGVPMG GI 2660712

1336 ELTIEFSKPLLPVGRAGVLLSEILHLLCKQ 1252206
1162 ELFREITKPLRPLGKAASLLLEILGLLCKS GI 2660712

1366 MSHKKVGALWREADLSWKDFLPEGEDVHNF 1252206
1192 MGPKKVGTLWREAGLSWKEFLPEGQDIGAF GI 2660712

1396 LLEQKLDFIESDSPCSSEALSKKELSAEEL 1252206
1222 VAEQKVEYTLGE---ESEAPGQRALPSEEL GI 2660712
```

FIGURE 2G

```
1426 YKRLEKLIIEDKANDEQIFDWVEANLDEIQ 1252206
1249 NRQLEKLLKEG-SSNQRVFDWIEANLSEQQ GI 2660712

1456 MSSPTFLRALMTAVCKAAIIADSSTFRVDT 1252206
1278 IVSNTLVRALMTAVCYSAIIFETP-LRVDV GI 2660712

1486 AVIKQRVPILLKYLDSDTEKELQALYALQA 1252206
1307 AVLKARAKLLQKYL-CDEQKELQALYALQA GI 2660712

1516 SIVKLDQPANLLRMFFDCLYDEEVISEDAF 1252206
1336 LVVTLEQPPNLLRMFFDALYDEDVVKEDAF GI 2660712

1546 YKWESSKDPAEQNGKGVALKSVTAFFTWLR 1252206
1366 YSWESSKDPAEQQGKGVALKSVTAFFKWLR GI 2660712

1576 EAEEEESEDN 1252206
1396 EAEEEESDHN GI 2660712
```

FIGURE 2H

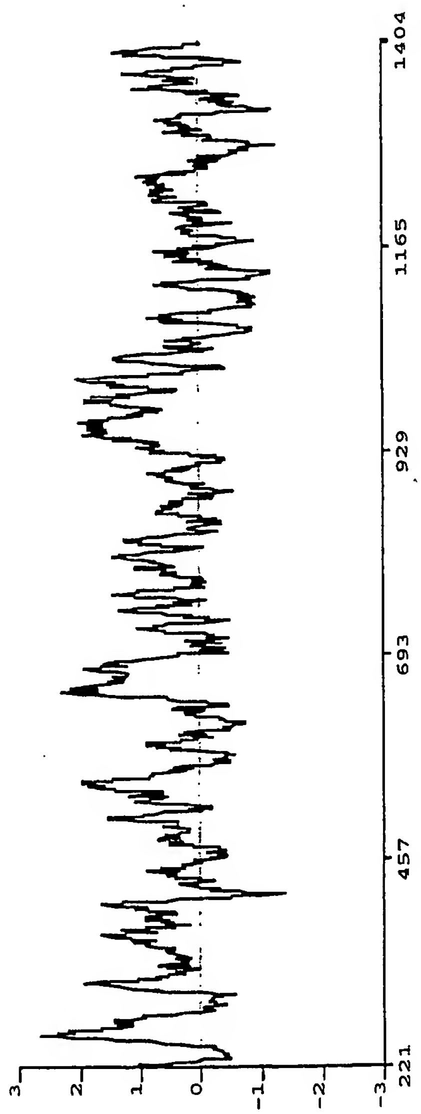


FIGURE 3A

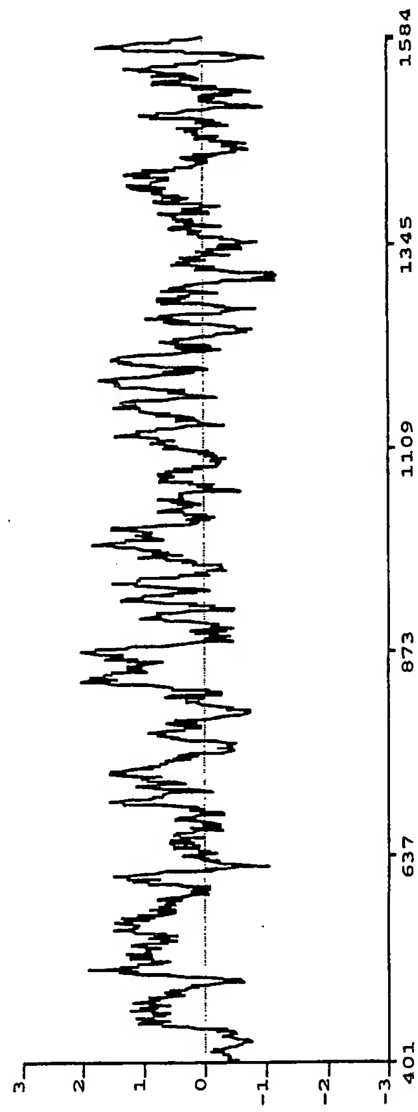


FIGURE 3B

```
1    M - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  2950994
1    L R R G A R N W R S M S T G E L T P Q S R L K E F S E L A R  GI 2440051

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  2950994
31   A L N L Y R M D H L G N Y T G H K S Y Y L T G Q L A T L E Q  GI 2440051

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  2950994
61   A I I Q Y A L Q A V T E H G F K L I S V P D I L P K E V I E  GI 2440051

2    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  2950994
91   S C G M R T E G E R T Q V Y K L D T G E C L S G T S E M A L  GI 2440051

2    - - - - - - - - - - - - - - - - - - - - - - - - F G V T G P  2950994
121  A G F F A N K L L S E D Q L P L K V T A V S R C Y R A E T S  GI 2440051

8    G L - - - - - - - - - - - - - - - - - - - - - - - E Q S S Q  2950994
151  G L Q E E K G I Y R V H Q F N K V E M F A I C T E E Q S E A  GI 2440051
```

FIGURE 4A

```
15   LLEEFLSLQMEILTELGLHFRVLDMPTQEL  2950994
181  ELEEFKNIEVDLFRRLGLNFRLLDMPPCEL  GI 2440051

45   GLPAYRKFDIEAWMPGRGRFGEVTSASNCT  2950994
211  GAPAYQKYDIEAWMPGRQMWGEISSCSNCT  GI 2440051

75   DFQSRRLHIMFQTEA-GELQFAHTVNATAC  2950994
241  DYQARRLGIRYRRSADGQILHAHTINGTAT  GI 2440051

104  AVPRLLIALLESNQQKDGSVLVPPALQSYL  2950994
271  AIPRLLIALLESYQ-KEDGIEIPAVLRPFM  GI 2440051

134  GTDR-ITAPTHVP----LQYIGPNQPRKPG  2950994
300  DNQELITRNKRIPETKLVKFIKA         GI 2440051

159  LPGQPAVS                        2950994
322                                  GI 2440051
```

FIGURE 4B

```
1    M C S L A S G A T G G R G A V E N E E D L P E L S D S G D E  3461657
1    M - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  GI 1808648

31   A A W E D E D D A D L P H G K Q Q T P C L F C N R L F T S A  3461657
2    - - - - - - - - - - - - - - - - E V S C - - - - - - - - - -  GI 1808648

61   E E T F S H C K S E H Q F N I D S M V H K H G L E F Y G Y I  3461657
6    - - - - G Q A E S S E K P N A E D M T S K - - - - - - - - -  GI 1808648

91   K L I N F I R L K N P T V E Y M N S I Y N P V P W E K E E Y  3461657
23   - - - - - - - - - - - - - D Y - - - - - - - - - - - - - - -  GI 1808648

121  L K P V L E D D L L L Q F D V E D L Y E P V S V P F S Y P N  3461657
25   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  GI 1808648

151  G L S E N T S V V E K L K H M E A R A L S A E A A L A R A R  3461657
25   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  GI 1808648
```

FIGURE 5A

```
181  E D L Q K M K Q F A Q D F V M H T D V R T C S S S T S V I A  3461657
25   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  GI 1808648

211  D L Q E D E D G V Y F S S Y G H Y G I H E E M L K D K I R T  3461657
25   - - - - - - - - - Y F D S Y A H F G I H E E M L K D E V R T  GI 1808648

241  E S Y R D F I Y Q N P H I F K D K V V L D V G C G T G I L S  3461657
46   L T Y R N S M F H N R H L F K D K V V L D V G S G T G I L C  GI 1808648

271  M F A A K A G A K K V L G V D Q S E I L Y Q A M D I I R L N  3461657
76   M F A A K A G A R K V I G I V C S S I S D Y A V K I V K A N  GI 1808648

301  K L E D T I T L I K G K I E E V H L P V E K V D V I I S E W  3461657
106  K L D H V V T I I K G K V E E V E L P V E K V D I I I S E W  GI 1808648

331  M G Y F L L F E S M L D S V L Y A K N K Y L A K G G S V Y P  3461657
136  M G Y C L F Y E S M L N T V L Y A R D K W L A P D G L I F P  GI 1808648
```

FIGURE 5B

```
361 D I C T I S L V A V S D V N K H A D R I A F W D D V Y G F K 3461657
166 D R A T L Y V T A I E D R Q Y K D Y K I H W W E N V Y G F D GI 1808648

391 M S C M K K A V I P E A V V E V L D P K T L I S E P C G I K 3461657
196 M S C I K D V A I K E P L V D V V D P K Q L V T N A C L I K GI 1808648

421 H I D C H T T S I S D L E F S S D F T L K I T R T S M C T A 3461657
226 E V D I Y T V K V E D L T F T S P F C L Q V K R N D Y V H A GI 1808648

451 I A G Y F D I Y F E K N C H N R V V F S T G P Q S T K T H W 3461657
256 L V A Y F N I E F T R - C H K R T G F S T S P E S P Y T H W GI 1808648

481 K Q T V F L L E K P F S V K A G E A L K G K V T V H K N K K 3461657
285 K Q T V F Y M E D Y L T V K T G E E I F G T I G M R P N A K GI 1808648

511 D P R S L T V T L T L N - - - - - - - - - N S T Q T Y G L Q 3461657
315 N N R D L D F T I D L D F K G Q L C E L S C S T D Y R M R GI 1808648
```

FIGURE 5C

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
TANG, Y. Tom
CORLEY, Neil C.
GUEGLER, Karl J.
GORGONE, Gina A.
PATTERSON, Chandra
HILLMAN, Jennifer L.
BAUGHN, Mariah R.
LAL, Preeti
AZIMZAI, Yalda
YUE, Henry
YANG, Junming

<120> RNA-ASSOCIATED PROTEINS

<130> PF-0600 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/156,039; unassigned; 09/158,720; unassigned; 09/186,815; unassigned; 60/128,660
<151> 1998-09-17; 1998-09-17; 1998-09-22; 1998-09-22; 1998-11-04; 1998-11-04; 1999-04-08

<160> 38

<170> PERL Program
<210> 1
<211> 216
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 399781CD1

<400> 1
```
Met Ser Arg Tyr Leu Arg Pro Pro Asn Thr Ser Leu Phe Val Arg
  1               5                  10                  15
Asn Val Ala Asp Asp Thr Arg Ser Glu Asp Leu Arg Arg Glu Phe
                 20                  25                  30
Gly Arg Tyr Gly Pro Ile Val Asp Val Tyr Val Pro Leu Asp Phe
                 35                  40                  45
Tyr Thr Arg Arg Pro Arg Gly Phe Ala Tyr Val Gln Phe Glu Asp
                 50                  55                  60
Val Arg Asp Ala Glu Asp Ala Leu His Asn Leu Asp Arg Lys Trp
                 65                  70                  75
Ile Cys Gly Arg Gln Ile Glu Ile Gln Phe Ala Gln Gly Asp Arg
                 80                  85                  90
Lys Thr Pro Asn Gln Met Lys Ala Lys Glu Gly Arg Asn Val Tyr
                 95                 100                 105
Ser Ser Ser Arg Tyr Asp Asp Tyr Asp Arg Tyr Arg Arg Ser Arg
                110                 115                 120
```

Ser Arg Ser Tyr Glu Arg Arg Arg Ser Arg Ser Arg Ser Phe Asp
125 130 135
Tyr Asn Tyr Arg Arg Ser Tyr Ser Pro Arg Asn Ser Arg Pro Thr
140 145 150
Gly Arg Pro Arg Arg Arg Glu Ala Ile Pro Thr Met Ile Asp Gln
155 160 165
Thr Ala Ala Gly Ile Pro Ser Thr Val Leu Leu Thr Thr Leu Gln
170 175 180
Glu Arg Ser Glu Ser Gly Lys Arg Thr Lys Glu Gly Gln Phe Lys
185 190 195
Arg Pro Lys Gly Gly Trp Lys Val Leu Gln Tyr Glu Tyr Cys Thr
200 205 210
Asn Ile Leu Thr Leu Val
215

<210> 2
<211> 1584
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1252206CD1

<400> 2
Met Asn Ser Gln Pro Gln Thr Arg Ser Pro Phe Phe Gln Arg Pro
1 5 10 15
Gln Ile Gln Pro Pro Arg Ala Thr Ile Pro Asn Ser Ser Pro Ser
20 25 30
Ile Arg Pro Gly Ala Gln Thr Pro Thr Ala Val Tyr Gln Ala Asn
35 40 45
Gln His Ile Met Met Val Asn His Leu Pro Met Pro Tyr Pro Val
50 55 60
Pro Gln Gly Pro Gln Tyr Cys Ile Pro Gln Tyr Arg His Ser Gly
65 70 75
Pro Pro Tyr Val Gly Pro Pro Gln Lys Tyr Pro Val Gln Pro Pro
80 85 90
Gly Pro Gly Pro Phe Tyr Pro Gly Pro Gly Pro Gly Asp Phe Pro
95 100 105
Asn Ala Tyr Gly Thr Pro Phe Tyr Pro Ser Gln Pro Val Tyr Gln
110 115 120
Ser Ala Pro Ile Ile Val Pro Thr Gln Gln Gln Pro Pro Pro Ala
125 130 135
Lys Arg Glu Lys Lys Thr Ile Arg Ile Arg Asp Pro Asn Gln Gly
140 145 150
Gly Lys Asp Ile Thr Glu Glu Ile Met Ser Gly Gly Gly Ser Arg
155 160 165
Asn Pro Thr Pro Pro Ile Gly Arg Pro Thr Ser Thr Pro Thr Pro
170 175 180
Pro Gln Leu Pro Ser Gln Val Pro Glu His Ser Pro Val Val Tyr
185 190 195
Gly Thr Val Glu Ser Ala His Leu Ala Ala Ser Thr Pro Val Thr
200 205 210
Ala Ala Ser Asp Gln Lys Gln Glu Glu Lys Pro Lys Pro Asp Pro
215 220 225
Val Leu Lys Ser Pro Ser Pro Val Leu Arg Leu Val Leu Ser Gly

230 235 240
Glu Lys Lys Glu Gln Glu Gly Gln Thr Ser Glu Thr Thr Ala Ile
245 250 255
Val Ser Ile Ala Glu Leu Pro Leu Pro Pro Ser Pro Thr Thr Val
260 265 270
Ser Ser Val Ala Arg Ser Thr Ile Ala Ala Pro Thr Ser Ser Ala
275 280 285
Leu Ser Ser Gln Pro Ile Phe Thr Thr Ala Ile Asp Asp Arg Cys
290 295 300
Glu Leu Ser Ser Pro Arg Glu Asp Thr Ile Pro Ile Pro Ser Leu
305 310 315
Thr Ser Cys Thr Glu Thr Ser Asp Pro Leu Pro Thr Asn Glu Asn
320 325 330
Asp Asp Asp Ile Cys Lys Lys Pro Cys Ser Val Ala Pro Asn Asp
335 340 345
Ile Pro Leu Val Ser Ser Thr Asn Leu Ile Asn Glu Ile Asn Gly
350 355 360
Val Ser Glu Lys Leu Ser Ala Thr Glu Ser Ile Val Glu Ile Val
365 370 375
Lys Gln Glu Val Leu Pro Leu Thr Leu Glu Leu Glu Ile Leu Glu
380 385 390
Asn Pro Pro Glu Glu Met Lys Leu Glu Cys Ile Pro Ala Pro Ile
395 400 405
Thr Pro Ser Thr Val Pro Ser Phe Pro Pro Thr Pro Pro Thr Pro
410 415 420
Pro Ala Ser Pro Pro His Thr Pro Val Ile Val Pro Ala Ala Ala
425 430 435
Thr Thr Val Ser Ser Pro Ser Ala Ala Ile Thr Val Gln Arg Val
440 445 450
Leu Glu Glu Asp Glu Ser Ile Arg Thr Cys Leu Ser Glu Asp Ala
455 460 465
Lys Glu Ile Gln Asn Lys Ile Glu Val Glu Ala Asp Gly Gln Thr
470 475 480
Glu Glu Ile Leu Asp Ser Gln Asn Leu Asn Ser Arg Arg Ser Pro
485 490 495
Val Pro Ala Gln Ile Ala Ile Thr Val Pro Lys Thr Trp Lys Lys
500 505 510
Pro Lys Asp Arg Thr Arg Thr Thr Glu Glu Met Leu Glu Ala Glu
515 520 525
Leu Glu Leu Lys Ala Glu Glu Glu Leu Ser Ile Asp Lys Val Leu
530 535 540
Glu Ser Glu Gln Asp Lys Met Ser Gln Gly Phe His Pro Glu Arg
545 550 555
Asp Pro Ser Asp Leu Lys Lys Val Lys Ala Val Glu Glu Asn Gly
560 565 570
Glu Glu Ala Glu Pro Val Arg Asn Gly Ala Glu Ser Val Ser Glu
575 580 585
Gly Glu Gly Ile Asp Ala Asn Ser Gly Ser Thr Asp Ser Ser Gly
590 595 600
Asp Gly Val Thr Phe Pro Phe Lys Pro Glu Ser Trp Lys Pro Thr
605 610 615
Asp Thr Glu Gly Lys Lys Gln Tyr Asp Arg Glu Phe Leu Leu Asp
620 625 630
Phe Gln Phe Met Pro Ala Cys Ile Gln Lys Pro Glu Gly Leu Pro
635 640 645
Pro Ile Ser Asp Val Val Leu Asp Lys Ile Asn Gln Pro Lys Leu
650 655 660

Pro Met Arg Thr Leu Asp Pro Arg Ile Leu Pro Arg Gly Pro Asp
665 670 675
Phe Thr Pro Ala Phe Ala Asp Phe Gly Arg Gln Thr Pro Gly Gly
680 685 690
Arg Gly Val Pro Leu Leu Asn Val Gly Ser Arg Arg Ser Gln Pro
695 700 705
Gly Gln Arg Arg Glu Pro Arg Lys Ile Ile Thr Val Ser Val Lys
710 715 720
Glu Asp Val His Leu Lys Lys Ala Glu Asn Ala Trp Lys Pro Ser
725 730 735
Gln Lys Arg Asp Ser Gln Ala Asp Asp Pro Glu Asn Ile Lys Thr
740 745 750
Gln Glu Leu Phe Arg Lys Val Arg Ser Ile Leu Asn Lys Leu Thr
755 760 765
Pro Gln Met Phe Asn Gln Leu Met Lys Gln Val Ser Gly Leu Thr
770 775 780
Val Asp Thr Glu Glu Arg Leu Lys Gly Val Ile Asp Leu Val Phe
785 790 795
Glu Lys Ala Ile Asp Glu Pro Ser Phe Ser Val Ala Tyr Ala Asn
800 805 810
Met Cys Arg Cys Leu Val Thr Leu Lys Val Pro Met Ala Asp Lys
815 820 825
Pro Gly Asn Thr Val Asn Phe Arg Lys Leu Leu Leu Asn Arg Cys
830 835 840
Gln Lys Glu Phe Glu Lys Asp Lys Ala Asp Asp Asp Val Phe Glu
845 850 855
Lys Lys Gln Lys Glu Leu Glu Ala Ala Ser Ala Pro Glu Glu Arg
860 865 870
Thr Arg Leu His Asp Glu Leu Glu Glu Ala Lys Asp Lys Ala Arg
875 880 885
Arg Arg Ser Ile Gly Asn Ile Lys Phe Ile Gly Glu Leu Phe Lys
890 895 900
Leu Lys Met Leu Thr Glu Ala Ile Met His Asp Cys Val Val Lys
905 910 915
Leu Leu Lys Asn His Asp Glu Glu Ser Leu Glu Cys Leu Cys Arg
920 925 930
Leu Leu Thr Thr Ile Gly Lys Asp Leu Asp Phe Glu Lys Ala Lys
935 940 945
Pro Arg Met Asp Gln Tyr Phe Asn Gln Met Glu Lys Ile Val Lys
950 955 960
Glu Lys Lys Thr Ser Ser Arg Ile Arg Phe Met Leu Gln Asp Val
965 970 975
Ile Asp Leu Arg Leu Cys Asn Trp Val Ser Arg Arg Ala Asp Gln
980 985 990
Gly Pro Lys Thr Ile Glu Gln Ile His Lys Glu Ala Lys Ile Glu
995 1000 1005
Glu Gln Glu Glu Gln Arg Lys Val Gln Gln Leu Met Thr Lys Glu
1010 1015 1020
Lys Arg Arg Pro Gly Val Gln Arg Val Asp Glu Gly Gly Trp Asn
1025 1030 1035
Thr Val Gln Gly Ala Lys Asn Ser Arg Val Leu Asp Pro Ser Lys
1040 1045 1050
Phe Leu Lys Ile Thr Lys Pro Thr Ile Asp Glu Lys Ile Gln Leu
1055 1060 1065
Val Pro Lys Ala Gln Leu Gly Ser Trp Gly Lys Gly Ser Ser Gly
1070 1075 1080
Gly Ala Lys Ala Ser Glu Thr Asp Ala Leu Arg Ser Ser Ala Ser

1085 1090 1095
Ser Leu Asn Arg Phe Ser Ala Leu Gln Pro Pro Ala Pro Ser Gly
1100 1105 1110
Ser Thr Pro Ser Thr Pro Val Glu Phe Asp Ser Arg Arg Thr Leu
1115 1120 1125
Thr Ser Arg Gly Ser Met Gly Arg Glu Lys Asn Asp Lys Pro Leu
1130 1135 1140
Pro Ser Ala Thr Ala Arg Pro Asn Thr Phe Met Arg Gly Gly Ser
1145 1150 1155
Ser Lys Asp Leu Leu Asp Asn Gln Ser Gln Glu Glu Gln Arg Arg
1160 1165 1170
Glu Met Leu Glu Thr Val Lys Gln Leu Thr Gly Gly Val Asp Val
1175 1180 1185
Glu Arg Asn Ser Thr Glu Ala Glu Arg Asn Lys Thr Arg Glu Ser
1190 1195 1200
Ala Lys Pro Glu Ile Ser Ala Met Ser Ala His Asp Lys Ala Ala
1205 1210 1215
Leu Ser Glu Glu Glu Leu Glu Arg Lys Ser Lys Ser Ile Ile Asp
1220 1225 1230
Glu Phe Leu His Ile Asn Asp Phe Lys Glu Ala Met Gln Cys Val
1235 1240 1245
Glu Glu Leu Asn Ala Gln Gly Leu Leu His Val Phe Val Arg Val
1250 1255 1260
Gly Val Glu Ser Thr Leu Glu Arg Ser Gln Ile Thr Arg Asp His
1265 1270 1275
Met Gly Gln Leu Leu Tyr Gln Leu Val Gln Ser Glu Lys Leu Ser
1280 1285 1290
Lys Gln Asp Phe Phe Lys Gly Phe Ser Glu Thr Leu Glu Leu Ala
1295 1300 1305
Asp Asp Met Ala Ile Asp Ile Pro His Ile Trp Leu Tyr Leu Ala
1310 1315 1320
Glu Leu Val Thr Pro Met Leu Lys Glu Gly Gly Ile Ser Met Arg
1325 1330 1335
Glu Leu Thr Ile Glu Phe Ser Lys Pro Leu Leu Pro Val Gly Arg
1340 1345 1350
Ala Gly Val Leu Leu Ser Glu Ile Leu His Leu Leu Cys Lys Gln
1355 1360 1365
Met Ser His Lys Lys Val Gly Ala Leu Trp Arg Glu Ala Asp Leu
1370 1375 1380
Ser Trp Lys Asp Phe Leu Pro Glu Gly Glu Asp Val His Asn Phe
1385 1390 1395
Leu Leu Glu Gln Lys Leu Asp Phe Ile Glu Ser Asp Ser Pro Cys
1400 1405 1410
Ser Ser Glu Ala Leu Ser Lys Lys Glu Leu Ser Ala Glu Glu Leu
1415 1420 1425
Tyr Lys Arg Leu Glu Lys Leu Ile Ile Glu Asp Lys Ala Asn Asp
1430 1435 1440
Glu Gln Ile Phe Asp Trp Val Glu Ala Asn Leu Asp Glu Ile Gln
1445 1450 1455
Met Ser Ser Pro Thr Phe Leu Arg Ala Leu Met Thr Ala Val Cys
1460 1465 1470
Lys Ala Ala Ile Ile Ala Asp Ser Ser Thr Phe Arg Val Asp Thr
1475 1480 1485
Ala Val Ile Lys Gln Arg Val Pro Ile Leu Leu Lys Tyr Leu Asp
1490 1495 1500
Ser Asp Thr Glu Lys Glu Leu Gln Ala Leu Tyr Ala Leu Gln Ala
1505 1510 1515

Ser Ile Val Lys Leu Asp Gln Pro Ala Asn Leu Leu Arg Met Phe
1520 1525 1530
Phe Asp Cys Leu Tyr Asp Glu Glu Val Ile Ser Glu Asp Ala Phe
1535 1540 1545
Tyr Lys Trp Glu Ser Ser Lys Asp Pro Ala Glu Gln Asn Gly Lys
1550 1555 1560
Gly Val Ala Leu Lys Ser Val Thr Ala Phe Phe Thr Trp Leu Arg
1565 1570 1575
Glu Ala Glu Glu Glu Ser Glu Asp Asn
1580

<210> 3
<211> 166
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2950994CD1

<400> 3
Met Phe Gly Val Thr Gly Pro Gly Leu Glu Gln Ser Ser Gln Leu
1 5 10 15
Leu Glu Glu Phe Leu Ser Leu Gln Met Glu Ile Leu Thr Glu Leu
20 25 30
Gly Leu His Phe Arg Val Leu Asp Met Pro Thr Gln Glu Leu Gly
35 40 45
Leu Pro Ala Tyr Arg Lys Phe Asp Ile Glu Ala Trp Met Pro Gly
50 55 60
Arg Gly Arg Phe Gly Glu Val Thr Ser Ala Ser Asn Cys Thr Asp
65 70 75
Phe Gln Ser Arg Arg Leu His Ile Met Phe Gln Thr Glu Ala Gly
80 85 90
Glu Leu Gln Phe Ala His Thr Val Asn Ala Thr Ala Cys Ala Val
95 100 105
Pro Arg Leu Leu Ile Ala Leu Leu Glu Ser Asn Gln Gln Lys Asp
110 115 120
Gly Ser Val Leu Val Pro Pro Ala Leu Gln Ser Tyr Leu Gly Thr
125 130 135
Asp Arg Ile Thr Ala Pro Thr His Val Pro Leu Gln Tyr Ile Gly
140 145 150
Pro Asn Gln Pro Arg Lys Pro Gly Leu Pro Gly Gln Pro Ala Val
155 160 165
Ser

<210> 4
<211> 531
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3461657CD1

<400> 4

Met Cys Ser Leu Ala Ser Gly Ala Thr Gly Gly Arg Gly Ala Val
1 5 10 15
Glu Asn Glu Glu Asp Leu Pro Glu Leu Ser Asp Ser Gly Asp Glu
20 25 30
Ala Ala Trp Glu Asp Glu Asp Asp Ala Asp Leu Pro His Gly Lys
35 40 45
Gln Gln Thr Pro Cys Leu Phe Cys Asn Arg Leu Phe Thr Ser Ala
50 55 60
Glu Glu Thr Phe Ser His Cys Lys Ser Glu His Gln Phe Asn Ile
65 70 75
Asp Ser Met Val His Lys His Gly Leu Glu Phe Tyr Gly Tyr Ile
80 85 90
Lys Leu Ile Asn Phe Ile Arg Leu Lys Asn Pro Thr Val Glu Tyr
95 100 105
Met Asn Ser Ile Tyr Asn Pro Val Pro Trp Glu Lys Glu Glu Tyr
110 115 120
Leu Lys Pro Val Leu Glu Asp Asp Leu Leu Leu Gln Phe Asp Val
125 130 135
Glu Asp Leu Tyr Glu Pro Val Ser Val Pro Phe Ser Tyr Pro Asn
140 145 150
Gly Leu Ser Glu Asn Thr Ser Val Val Glu Lys Leu Lys His Met
155 160 165
Glu Ala Arg Ala Leu Ser Ala Glu Ala Ala Leu Ala Arg Ala Arg
170 175 180
Glu Asp Leu Gln Lys Met Lys Gln Phe Ala Gln Asp Phe Val Met
185 190 195
His Thr Asp Val Arg Thr Cys Ser Ser Ser Thr Ser Val Ile Ala
200 205 210
Asp Leu Gln Glu Asp Glu Asp Gly Val Tyr Phe Ser Ser Tyr Gly
215 220 225
His Tyr Gly Ile His Glu Glu Met Leu Lys Asp Lys Ile Arg Thr
230 235 240
Glu Ser Tyr Arg Asp Phe Ile Tyr Gln Asn Pro His Ile Phe Lys
245 250 255
Asp Lys Val Val Leu Asp Val Gly Cys Gly Thr Gly Ile Leu Ser
260 265 270
Met Phe Ala Ala Lys Ala Gly Ala Lys Lys Val Leu Gly Val Asp
275 280 285
Gln Ser Glu Ile Leu Tyr Gln Ala Met Asp Ile Ile Arg Leu Asn
290 295 300
Lys Leu Glu Asp Thr Ile Thr Leu Ile Lys Gly Lys Ile Glu Glu
305 310 315
Val His Leu Pro Val Glu Lys Val Asp Val Ile Ile Ser Glu Trp
320 325 330
Met Gly Tyr Phe Leu Leu Phe Glu Ser Met Leu Asp Ser Val Leu
335 340 345
Tyr Ala Lys Asn Lys Tyr Leu Ala Lys Gly Gly Ser Val Tyr Pro
350 355 360
Asp Ile Cys Thr Ile Ser Leu Val Ala Val Ser Asp Val Asn Lys
365 370 375
His Ala Asp Arg Ile Ala Phe Trp Asp Asp Val Tyr Gly Phe Lys
380 385 390
Met Ser Cys Met Lys Lys Ala Val Ile Pro Glu Ala Val Val Glu
395 400 405
Val Leu Asp Pro Lys Thr Leu Ile Ser Glu Pro Cys Gly Ile Lys
410 415 420
His Ile Asp Cys His Thr Thr Ser Ile Ser Asp Leu Glu Phe Ser

```
                425                           430                           435
Ser Asp Phe Thr Leu Lys Ile Thr Arg Thr Ser Met Cys Thr Ala
                440                           445                           450
Ile Ala Gly Tyr Phe Asp Ile Tyr Phe Glu Lys Asn Cys His Asn
                455                           460                           465
Arg Val Val Phe Ser Thr Gly Pro Gln Ser Thr Lys Thr His Trp
                470                           475                           480
Lys Gln Thr Val Phe Leu Leu Glu Lys Pro Phe Ser Val Lys Ala
                485                           490                           495
Gly Glu Ala Leu Lys Gly Lys Val Thr Val His Lys Asn Lys Lys
                500                           505                           510
Asp Pro Arg Ser Leu Thr Val Thr Leu Thr Leu Asn Asn Ser Thr
                515                           520                           525
Gln Thr Tyr Gly Leu Gln
                530
```

<210> 5
<211> 148
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 053076CD1

<400> 5

```
Met Ala Ser Val Val Leu Ala Leu Arg Thr Arg Thr Ala Val Thr
 1               5                   10                  15
Ser Leu Leu Ser Pro Thr Pro Ala Thr Ala Leu Ala Val Arg Tyr
                20                  25                  30
Ala Ser Lys Lys Ser Gly Gly Ser Ser Lys Asn Leu Gly Gly Lys
                35                  40                  45
Ser Ser Gly Arg Arg Gln Gly Ile Lys Lys Met Glu Gly His Tyr
                50                  55                  60
Val His Ala Gly Asn Ile Ile Ala Thr Gln Arg His Phe Arg Trp
                65                  70                  75
His Pro Gly Ala His Val Gly Val Gly Lys Asn Lys Cys Leu Tyr
                80                  85                  90
Ala Leu Glu Glu Gly Ile Val Arg Tyr Thr Lys Glu Val Tyr Val
                95                  100                 105
Pro His Pro Arg Asn Thr Glu Ala Val Asp Leu Ile Thr Arg Leu
                110                 115                 120
Pro Lys Gly Ala Val Leu Tyr Lys Thr Phe Val His Val Val Pro
                125                 130                 135
Ala Lys Pro Glu Gly Thr Phe Lys Leu Val Ala Met Leu
                140                 145
```

<210> 6
<211> 317
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1292379CD1

<400> 6

```
Met Met Ser Phe His Ser Asn Arg Pro Ser Lys Arg Phe Cys Ile
  1               5                  10                  15
Phe Lys Lys His Ser Glu Asn Leu Arg Gly Ile Thr Leu Val Cys
                20                  25                  30
Leu Asn Cys Asp Phe Leu Ser Asp Val Ser Gly Leu Asp Asn Met
                35                  40                  45
Ala Thr His Leu Ser Gln His Lys Thr His Thr Cys Gln Val Val
                50                  55                  60
Met Gln Lys Val Ser Val Cys Ile Pro Thr Ser Glu His Leu Ser
                65                  70                  75
Glu Leu Lys Lys Glu Ala Pro Ala Lys Glu Gln Glu Pro Val Ser
                80                  85                  90
Lys Glu Ile Ala Arg Pro Asn Met Ala Glu Arg Glu Thr Glu Thr
                95                 100                 105
Ser Asn Ser Glu Ser Lys Gln Asp Lys Ala Ala Ser Ser Lys Glu
               110                 115                 120
Lys Asn Gly Cys Asn Ala Asn Ser Phe Glu Gly Ser Ser Thr Thr
               125                 130                 135
Lys Ser Glu Glu Ser Ile Thr Val Ser Asp Lys Glu Asn Glu Thr
               140                 145                 150
Cys Leu Ala Asp Gln Glu Thr Gly Ser Lys Asn Ile Val Ser Cys
               155                 160                 165
Asp Ser Asn Ile Gly Ala Asp Lys Val Glu Lys Lys Lys Gln Ile
               170                 175                 180
Gln His Val Cys Gln Glu Met Glu Leu Lys Met Cys Gln Ser Ser
               185                 190                 195
Glu Asn Ile Ile Leu Ser Asp Gln Ile Lys Asp His Asn Ser Ser
               200                 205                 210
Glu Ala Arg Phe Ser Ser Lys Asn Ile Lys Asp Leu Arg Leu Ala
               215                 220                 225
Ser Asp Asn Val Ser Ile Asp Gln Phe Leu Arg Lys Arg His Glu
               230                 235                 240
Pro Glu Ser Val Ser Ser Asp Val Ser Glu Gln Gly Ser Ile His
               245                 250                 255
Leu Glu Pro Leu Thr Pro Ser Glu Val Leu Glu Tyr Glu Ala Thr
               260                 265                 270
Glu Ile Leu Gln Lys Gly Ser Gly Asp Pro Ser Ala Lys Thr Asp
               275                 280                 285
Glu Val Val Ser Asp Gln Thr Asp Asp Ile Pro Gly Gly Asn Asn
               290                 295                 300
Pro Ser Thr Thr Glu Ala Thr Val Asp Leu Glu Asp Glu Lys Glu
               305                 310                 315
Arg Ser
```

<210> 7
<211> 278
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1437783CD1

<400> 7
Met Ala Ala Leu Phe Leu Lys Arg Leu Thr Leu Gln Thr Val Lys
1 5 10 15
Ser Glu Asn Ser Cys Ile Arg Cys Phe Gly Lys His Ile Leu Gln
20 25 30
Lys Thr Ala Pro Ala Gln Leu Ser Pro Ile Ala Ser Ala Pro Arg
35 40 45
Leu Ser Phe Leu Ile His Ala Lys Ala Phe Ser Thr Ala Glu Asp
50 55 60
Thr Gln Asn Glu Gly Lys Lys Thr Lys Lys Asn Lys Thr Ala Phe
65 70 75
Ser Asn Val Gly Arg Lys Ile Ser Gln Arg Val Ile His Leu Phe
80 85 90
Asp Glu Lys Gly Asn Asp Leu Gly Asn Met His Arg Ala Asn Val
95 100 105
Ile Arg Leu Met Asp Glu Arg Asp Leu Arg Leu Val Gln Arg Asn
110 115 120
Thr Ser Thr Glu Pro Ala Glu Tyr Gln Leu Met Thr Gly Leu Gln
125 130 135
Ile Leu Gln Glu Arg Gln Arg Leu Arg Glu Met Glu Lys Ala Asn
140 145 150
Pro Lys Thr Gly Pro Thr Leu Arg Lys Glu Leu Ile Leu Ser Ser
155 160 165
Asn Ile Gly Gln His Asp Leu Asp Thr Lys Thr Lys Gln Ile Gln
170 175 180
Gln Trp Ile Lys Lys Lys His Leu Val Gln Ile Thr Ile Lys Lys
185 190 195
Gly Lys Asn Val Asp Val Ser Glu Asn Glu Met Glu Glu Ile Phe
200 205 210
His Gln Ile Leu Gln Thr Met Pro Gly Ile Ala Thr Phe Ser Ser
215 220 225
Arg Pro Gln Ala Val Gln Gly Gly Lys Ala Leu Met Cys Val Leu
230 235 240
Arg Ala Leu Ser Lys Asn Glu Glu Lys Ala Tyr Lys Glu Thr Gln
245 250 255
Glu Thr Gln Glu Arg Asp Thr Leu Asn Lys Asp His Gly Asn Asp
260 265 270
Lys Glu Ser Asn Val Leu His Gln
275

<210> 8
<211> 586
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1557635CD1

<400> 8
Met Ser Ala Thr Val Val Asp Ala Val Asn Ala Ala Pro Leu Ser
1 5 10 15
Gly Ser Lys Glu Met Ser Leu Glu Glu Pro Lys Lys Met Thr Arg
20 25 30
Glu Asp Trp Arg Lys Lys Lys Glu Leu Glu Glu Gln Arg Lys Leu
35 40 45

```
Gly Asn Ala Pro Ala Glu Val Asp Glu Glu Gly Lys Asp Ile Asn
                 50                  55                  60
Pro His Ile Pro Gln Tyr Ile Ser Ser Val Pro Trp Tyr Ile Asp
                 65                  70                  75
Pro Ser Lys Arg Pro Thr Leu Lys His Gln Arg Pro Gln Pro Glu
                 80                  85                  90
Lys Gln Lys Gln Phe Ser Ser Ser Gly Glu Trp Tyr Lys Arg Gly
                 95                 100                 105
Val Lys Glu Asn Ser Ile Ile Thr Lys Tyr Arg Lys Gly Ala Cys
                110                 115                 120
Glu Asn Cys Gly Ala Met Thr His Lys Lys Lys Asp Cys Phe Glu
                125                 130                 135
Arg Pro Arg Arg Val Gly Ala Lys Phe Thr Gly Thr Asn Ile Ala
                140                 145                 150
Pro Asp Glu His Val Gln Pro Gln Leu Met Phe Asp Tyr Asp Gly
                155                 160                 165
Lys Arg Asp Arg Trp Asn Gly Tyr Asn Pro Glu Glu His Met Lys
                170                 175                 180
Ile Val Glu Glu Tyr Ala Lys Val Asp Leu Ala Lys Arg Thr Leu
                185                 190                 195
Lys Ala Gln Lys Leu Gln Glu Glu Leu Ala Ser Gly Lys Leu Val
                200                 205                 210
Glu Gln Ala Asn Ser Pro Lys His Gln Trp Gly Glu Glu Glu Pro
                215                 220                 225
Asn Ser Gln Thr Glu Lys Asp His Asn Ser Glu Asp Glu Asp Glu
                230                 235                 240
Asp Lys Tyr Ala Asp Asp Ile Asp Met Pro Gly Gln Asn Phe Asp
                245                 250                 255
Ser Lys Arg Arg Ile Thr Val Arg Asn Leu Arg Ile Arg Glu Asp
                260                 265                 270
Ile Ala Lys Tyr Leu Arg Asn Leu Asp Pro Asn Ser Ala Tyr Tyr
                275                 280                 285
Asp Pro Lys Thr Arg Ala Met Arg Glu Asn Pro Tyr Ala Asn Ala
                290                 295                 300
Gly Lys Asn Pro Asp Glu Val Ser Tyr Ala Gly Asp Asn Phe Val
                305                 310                 315
Arg Tyr Thr Gly Asp Thr Ile Ser Met Ala Gln Thr Gln Leu Phe
                320                 325                 330
Ala Trp Glu Ala Tyr Asp Lys Gly Ser Glu Val His Leu Gln Ala
                335                 340                 345
Asp Pro Thr Lys Leu Glu Leu Leu Tyr Lys Ser Phe Lys Val Lys
                350                 355                 360
Lys Glu Asp Phe Lys Glu Gln Gln Lys Glu Ser Ile Leu Glu Lys
                365                 370                 375
Tyr Gly Gly Gln Glu His Leu Asp Ala Pro Pro Ala Glu Leu Leu
                380                 385                 390
Leu Ala Gln Thr Glu Asp Tyr Val Glu Tyr Ser Arg His Gly Thr
                395                 400                 405
Val Ile Lys Gly Gln Glu Arg Ala Val Ala Cys Ser Lys Tyr Glu
                410                 415                 420
Glu Asp Val Lys Ile His Asn His Thr His Ile Trp Gly Ser Tyr
                425                 430                 435
Trp Lys Glu Gly Arg Trp Gly Tyr Lys Cys Cys His Ser Phe Phe
                440                 445                 450
Lys Tyr Ser Tyr Cys Thr Gly Glu Ala Gly Lys Glu Ile Val Asn
                455                 460                 465
Ser Glu Glu Cys Ile Ile Asn Glu Ile Thr Gly Glu Glu Ser Val
```

```
                    470                     475                     480
Lys Lys Pro Gln Thr Leu Met Glu Leu His Gln Glu Lys Leu Lys
                    485                     490                     495
Glu Glu Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys His Arg Lys
                    500                     505                     510
Ser Ser Ser Asp Ser Asp Asp Glu Glu Lys Lys His Glu Lys Leu
                    515                     520                     525
Lys Lys Ala Leu Asn Ala Glu Glu Ala Arg Leu Leu His Val Lys
                    530                     535                     540
Glu Thr Met Gln Ile Asp Glu Arg Lys Arg Pro Tyr Asn Ser Met
                    545                     550                     555
Tyr Glu Thr Arg Glu Pro Thr Glu Glu Glu Met Glu Ala Tyr Arg
                    560                     565                     570
Met Lys Arg Gln Arg Pro Asp Asp Pro Met Ala Ser Phe Leu Gly
                    575                     580                     585
Gln


&lt;210&gt; 9
&lt;211&gt; 384
&lt;212&gt; PRT
&lt;213&gt; Homo sapiens

&lt;220&gt;
&lt;221&gt; misc_feature
&lt;223&gt; Incyte ID No.: 2049352CD1

&lt;400&gt; 9
Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp
  1               5                  10                  15
Thr Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg
                 20                  25                  30
Asn Thr Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser
                 35                  40                  45
Arg Pro Pro Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser
                 50                  55                  60
Glu Leu Lys Tyr His Pro Glu Met Arg Phe Phe His Trp Phe Ser
                 65                  70                  75
Lys Trp Arg Lys Leu His Arg Asp Gln Glu Tyr Glu Val Thr Trp
                 80                  85                  90
Tyr Ile Ser Trp Ser Pro Cys Thr Lys Cys Thr Arg Asp Met Ala
                 95                 100                 105
Thr Phe Leu Ala Glu Asp Pro Lys Val Thr Leu Thr Ile Phe Val
                110                 115                 120
Ala Arg Leu Tyr Tyr Phe Trp Asp Pro Asp Tyr Gln Glu Ala Leu
                125                 130                 135
Arg Ser Leu Cys Gln Lys Arg Asp Gly Pro Arg Ala Thr Met Lys
                140                 145                 150
Ile Met Asn Tyr Asp Glu Phe Gln His Cys Trp Ser Lys Phe Val
                155                 160                 165
Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn Asn Leu Pro Lys
                170                 175                 180
Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile Leu Arg His
                185                 190                 195
Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro
                200                 205                 210
Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu
```

```
                215                           220                           225
Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
                230                           235                           240
Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly
                245                           250                           255
Arg His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys
                260                           265                           270
Leu Asp Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp
                275                           280                           285
Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser
                290                           295                           300
Lys Asn Lys His Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr
                305                           310                           315
Asp Asp Gln Gly Arg Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu
                320                           325                           330
Ala Gly Ala Lys Ile Ser Ile Leu Thr Tyr Ser Glu Phe Lys His
                335                           340                           345
Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro
                350                           355                           360
Trp Asp Gly Leu Glu Glu His Ser Gln Ala Leu Ser Gly Arg Leu
                365                           370                           375
Arg Gly Ile Leu Gln Asn Gln Gly Ser
                380


<210> 10
<211> 325
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2231663CD1

<400> 10
Met Ala Ala Ala Val Arg Cys Met Gly Arg Ala Leu Ile His His
  1             5                             10                            15
Gln Arg His Ser Leu Ser Lys Met Val Tyr Gln Thr Ser Leu Cys
                20                            25                            30
Ser Cys Ser Val Asn Ile Arg Val Pro Asn Arg His Phe Ala Ala
                35                            40                            45
Ala Thr Lys Ser Ala Lys Lys Thr Lys Lys Gly Ala Lys Glu Lys
                50                            55                            60
Thr Pro Asp Glu Lys Lys Asp Glu Ile Glu Lys Ile Lys Ala Tyr
                65                            70                            75
Pro Tyr Met Glu Gly Glu Pro Glu Asp Asp Val Tyr Leu Lys Arg
                80                            85                            90
Leu Tyr Pro Arg Gln Ile Tyr Glu Val Glu Lys Ala Val His Leu
                95                            100                           105
Leu Lys Lys Phe Gln Ile Leu Asp Phe Thr Ser Pro Lys Gln Ser
                110                           115                           120
Val Tyr Leu Asp Leu Thr Leu Asp Met Ala Leu Gly Lys Lys Lys
                125                           130                           135
Asn Val Glu Pro Phe Thr Ser Val Leu Ser Leu Pro Tyr Pro Phe
                140                           145                           150
Ala Ser Glu Ile Asn Lys Val Ala Val Phe Thr Glu Asn Ala Ser
                155                           160                           165
```

```
Glu Val Lys Ile Ala Glu Glu Asn Gly Ala Ala Phe Ala Gly Gly
                170                 175                 180
Thr Ser Leu Ile Gln Lys Ile Trp Asp Asp Glu Ile Val Ala Asp
                185                 190                 195
Phe Tyr Val Ala Val Pro Glu Ile Met Pro Glu Leu Asn Arg Leu
                200                 205                 210
Arg Lys Lys Leu Asn Lys Lys Tyr Pro Lys Leu Ser Arg Asn Ser
                215                 220                 225
Ile Gly Arg Asp Ile Pro Lys Met Leu Glu Leu Phe Lys Asn Gly
                230                 235                 240
His Glu Ile Lys Val Asp Glu Glu Arg Glu Asn Phe Leu Gln Thr
                245                 250                 255
Lys Ile Ala Thr Leu Asp Met Ser Ser Asp Gln Ile Ala Ala Asn
                260                 265                 270
Leu Gln Ala Val Ile Asn Glu Val Cys Arg His Arg Pro Leu Asn
                275                 280                 285
Leu Gly Pro Phe Val Val Arg Ala Phe Leu Arg Ser Ser Thr Ser
                290                 295                 300
Glu Gly Leu Leu Leu Lys Ile Asp Pro Leu Leu Pro Lys Glu Val
                305                 310                 315
Lys Asn Glu Glu Ser Glu Lys Glu Asp Ala
                320                 325


<210> 11
<211> 351
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2604449CD1

<400> 11
Met Gly Asp Pro Glu Arg Pro Glu Ala Ala Gly Leu Asp Gln Asp
  1               5                  10                  15
Glu Arg Ser Ser Ser Asp Thr Asn Glu Ser Glu Ile Lys Ser Asn
                 20                  25                  30
Glu Glu Pro Leu Leu Arg Lys Ser Ser Arg Arg Phe Val Ile Phe
                 35                  40                  45
Pro Ile Gln Tyr Pro Asp Ile Trp Lys Met Tyr Lys Gln Ala Gln
                 50                  55                  60
Ala Ser Phe Trp Thr Ala Glu Glu Val Asp Leu Ser Lys Asp Leu
                 65                  70                  75
Pro His Trp Asn Lys Leu Lys Ala Asp Glu Lys Tyr Phe Ile Ser
                 80                  85                  90
His Ile Leu Ala Phe Phe Ala Ala Ser Asp Gly Ile Val Asn Glu
                 95                 100                 105
Asn Leu Val Glu Arg Phe Ser Gln Glu Val Gln Val Pro Glu Ala
                110                 115                 120
Arg Cys Phe Tyr Gly Phe Gln Ile Leu Ile Glu Asn Val His Ser
                125                 130                 135
Glu Met Tyr Ser Leu Leu Ile Asp Thr Tyr Ile Arg Asp Pro Lys
                140                 145                 150
Lys Arg Glu Phe Leu Phe Asn Ala Ile Glu Thr Met Pro Tyr Val
                155                 160                 165
Lys Lys Lys Ala Asp Trp Ala Leu Arg Trp Ile Ala Asp Arg Lys
```

```
                    170                     175                     180
Ser Thr Phe Gly Glu Arg Val Val Ala Phe Ala Ala Val Glu Gly
                185                     190                     195
Val Phe Phe Ser Gly Ser Phe Ala Ala Ile Phe Trp Leu Lys Lys
                200                     205                     210
Arg Gly Leu Met Pro Gly Leu Thr Phe Ser Asn Glu Leu Ile Ser
                215                     220                     225
Arg Asp Glu Gly Leu His Cys Asp Phe Ala Cys Leu Met Phe Gln
                230                     235                     240
Tyr Leu Val Asn Lys Pro Ser Glu Glu Arg Val Arg Glu Ile Ile
                245                     250                     255
Val Asp Ala Val Lys Ile Glu Gln Glu Phe Leu Thr Glu Ala Leu
                260                     265                     270
Pro Val Gly Leu Ile Gly Met Asn Cys Ile Leu Met Lys Gln Tyr
                275                     280                     285
Ile Glu Phe Val Ala Asp Arg Leu Leu Val Glu Leu Gly Phe Ser
                290                     295                     300
Lys Val Phe Gln Ala Glu Asn Pro Phe Asp Phe Met Glu Asn Ile
                305                     310                     315
Ser Leu Glu Gly Lys Thr Asn Phe Phe Glu Lys Arg Val Ser Glu
                320                     325                     330
Tyr Gln Arg Phe Ala Val Met Ala Glu Thr Thr Asp Asn Val Phe
                335                     340                     345
Thr Leu Asp Ala Asp Phe
                350


<210> 12
<211> 681
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2604993CD1

<400> 12
Met Thr Ala Ser Pro Asp Tyr Leu Val Val Leu Phe Gly Ile Thr
  1               5                  10                  15
Ala Gly Ala Thr Gly Ala Lys Leu Gly Ser Asp Glu Lys Glu Leu
                 20                  25                  30
Ile Leu Leu Phe Trp Lys Val Val Asp Leu Ala Asn Lys Lys Val
                 35                  40                  45
Gly Gln Leu His Glu Val Leu Val Arg Pro Asp Gln Leu Glu Leu
                 50                  55                  60
Thr Glu Asp Cys Lys Glu Glu Thr Lys Ile Asp Val Glu Ser Leu
                 65                  70                  75
Ser Ser Ala Ser Gln Leu Asp Gln Ala Leu Arg Gln Phe Asn Gln
                 80                  85                  90
Ser Val Ser Asn Glu Leu Asn Ile Gly Val Gly Thr Ser Phe Cys
                 95                 100                 105
Leu Cys Thr Asp Gly Gln Leu His Val Arg Gln Ile Leu His Pro
                110                 115                 120
Glu Ala Ser Lys Lys Asn Val Leu Leu Pro Glu Cys Phe Tyr Ser
                125                 130                 135
Phe Phe Asp Leu Arg Lys Glu Phe Lys Lys Cys Cys Pro Gly Ser
                140                 145                 150
```

```
Pro Asp Ile Asp Lys Leu Asp Val Ala Thr Met Thr Glu Tyr Leu
                155                 160                 165
Asn Phe Glu Lys Ser Ser Ser Val Ser Arg Tyr Gly Ala Ser Gln
                170                 175                 180
Val Glu Asp Met Gly Asn Ile Ile Leu Ala Met Ile Ser Glu Pro
                185                 190                 195
Tyr Asn His Arg Phe Ser Asp Pro Glu Arg Val Asn Tyr Lys Phe
                200                 205                 210
Glu Ser Gly Thr Cys Ser Lys Met Glu Leu Ile Asp Asp Asn Thr
                215                 220                 225
Val Val Arg Ala Arg Gly Leu Pro Trp Gln Ser Ser Asp Gln Asp
                230                 235                 240
Ile Ala Arg Phe Phe Lys Gly Leu Asn Ile Ala Lys Gly Gly Ala
                245                 250                 255
Ala Leu Cys Leu Asn Ala Gln Gly Arg Arg Asn Gly Glu Ala Leu
                260                 265                 270
Val Arg Phe Val Ser Glu Glu His Arg Asp Leu Ala Leu Gln Arg
                275                 280                 285
His Lys His His Met Gly Thr Arg Tyr Ile Glu Val Tyr Lys Ala
                290                 295                 300
Thr Gly Glu Asp Phe Leu Lys Ile Ala Gly Gly Thr Ser Asn Glu
                305                 310                 315
Val Ala Gln Phe Leu Ser Lys Glu Asn Gln Val Ile Val Arg Met
                320                 325                 330
Arg Gly Leu Pro Phe Thr Ala Thr Ala Glu Glu Val Val Ala Phe
                335                 340                 345
Phe Gly Gln His Cys Pro Ile Thr Gly Gly Lys Glu Gly Ile Leu
                350                 355                 360
Phe Val Thr Tyr Pro Asp Gly Arg Pro Thr Gly Asp Ala Phe Val
                365                 370                 375
Leu Phe Ala Cys Glu Glu Tyr Ala Gln Asn Ala Leu Arg Lys His
                380                 385                 390
Lys Asp Leu Leu Gly Lys Arg Tyr Ile Glu Leu Phe Arg Ser Thr
                395                 400                 405
Ala Ala Glu Val Gln Gln Val Leu Asn Arg Phe Ser Ser Ala Pro
                410                 415                 420
Leu Ile Pro Leu Pro Thr Pro Pro Ile Ile Pro Val Leu Pro Gln
                425                 430                 435
Gln Phe Val Pro Pro Thr Asn Val Arg Asp Cys Ile Arg Leu Arg
                440                 445                 450
Gly Leu Pro Tyr Ala Ala Thr Ile Glu Asp Ile Leu Asp Phe Leu
                455                 460                 465
Gly Glu Phe Ala Thr Asp Ile Arg Thr His Gly Val His Met Val
                470                 475                 480
Leu Asn His Gln Gly Arg Pro Ser Gly Asp Ala Phe Ile Gln Met
                485                 490                 495
Lys Ser Ala Asp Arg Ala Phe Met Ala Ala Gln Lys Cys His Lys
                500                 505                 510
Lys Asn Met Lys Asp Arg Tyr Val Glu Val Phe Gln Cys Ser Ala
                515                 520                 525
Glu Glu Met Asn Phe Val Leu Met Gly Gly Thr Leu Asn Arg Asn
                530                 535                 540
Gly Leu Ser Pro Pro Pro Cys Lys Leu Pro Cys Leu Ser Pro Pro
                545                 550                 555
Ser Tyr Thr Phe Pro Ala Pro Ala Ala Val Ile Pro Thr Glu Ala
                560                 565                 570
Ala Ile Tyr Gln Pro Ser Val Ile Leu Asn Pro Arg Ala Leu Gln
```

```
                575                          580                          585
Pro Ser Thr Ala Tyr Tyr Pro Ala Gly Thr Gln Leu Phe Met Asn
                590                          595                          600
Tyr Thr Ala Tyr Tyr Pro Ser Pro Pro Gly Ser Pro Asn Ser Leu
                605                          610                          615
Gly Tyr Phe Pro Thr Ala Ala Asn Leu Ser Gly Val Pro Pro Gln
                620                          625                          630
Pro Gly Thr Val Val Arg Met Gln Gly Leu Ala Tyr Asn Thr Gly
                635                          640                          645
Val Lys Glu Ile Leu Asn Phe Phe Gln Gly Tyr Gln Tyr Ala Thr
                650                          655                          660
Glu Asp Gly Leu Ile His Thr Asn Asp Gln Ala Arg Thr Leu Pro
                665                          670                          675
Lys Glu Trp Val Cys Ile
                680
```

<210> 13
<211> 408
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2879070CD1

<400> 13

```
Met Ser Ser Leu Val Glu Thr Phe Val Ser Lys Ala Ser Ala Leu
 1               5                   10                  15
Gln Arg Gln Gly Arg Ala Gly Arg Val Arg Asp Gly Phe Cys Phe
                20                   25                  30
Arg Met Tyr Thr Arg Glu Arg Phe Glu Gly Phe Met Asp Tyr Ser
                35                   40                  45
Val Pro Glu Ile Leu Arg Val Pro Leu Glu Glu Leu Cys Leu His
                50                   55                  60
Ile Met Lys Cys Asn Leu Gly Ser Pro Glu Asp Phe Leu Ser Lys
                65                   70                  75
Ala Leu Asp Pro Pro Gln Leu Gln Val Ile Ser Asn Ala Met Asn
                80                   85                  90
Leu Leu Arg Lys Ile Gly Ala Cys Glu Leu Asn Glu Pro Lys Leu
                95                  100                 105
Thr Pro Leu Gly Gln His Leu Ala Ala Leu Pro Val Asn Val Lys
               110                  115                 120
Ile Gly Lys Met Leu Ile Phe Gly Ala Ile Phe Gly Cys Leu Asp
               125                  130                 135
Pro Val Ala Thr Leu Ala Ala Val Met Thr Glu Lys Ser Pro Phe
               140                  145                 150
Thr Thr Pro Ile Gly Arg Lys Asp Glu Ala Asp Leu Ala Lys Ser
               155                  160                 165
Ala Leu Ala Met Ala Asp Ser Asp His Leu Thr Ile Tyr Asn Ala
               170                  175                 180
Tyr Leu Gly Trp Lys Lys Ala Arg Gln Glu Gly Gly Tyr Arg Ser
               185                  190                 195
Glu Ile Thr Tyr Cys Arg Arg Asn Phe Leu Asn Arg Thr Ser Leu
               200                  205                 210
Leu Thr Leu Glu Asp Val Lys Gln Glu Leu Ile Lys Leu Val Lys
               215                  220                 225
```

Ala Ala Gly Phe Ser Ser Ser Thr Thr Ser Thr Ser Trp Glu Gly
230 235 240
Asn Arg Ala Ser Gln Thr Leu Ser Phe Gln Glu Ile Ala Leu Leu
245 250 255
Lys Ala Val Leu Val Ala Gly Leu Tyr Asp Asn Val Gly Lys Ile
260 265 270
Ile Tyr Thr Lys Ser Val Asp Val Thr Glu Lys Leu Ala Cys Ile
275 280 285
Val Glu Thr Ala Gln Gly Lys Ala Gln Val His Pro Ser Ser Val
290 295 300
Asn Arg Asp Leu Gln Thr His Gly Trp Leu Leu Tyr Gln Glu Lys
305 310 315
Ile Arg Tyr Ala Arg Val Tyr Leu Arg Glu Thr Thr Leu Ile Thr
320 325 330
Pro Phe Pro Val Leu Leu Phe Gly Gly Asp Ile Glu Val Gln His
335 340 345
Arg Glu Arg Leu Leu Ser Ile Asp Gly Trp Ile Tyr Phe Gln Ala
350 355 360
Pro Val Lys Ile Ala Val Ile Phe Lys Gln Leu Arg Val Leu Ile
365 370 375
Asp Ser Val Leu Arg Lys Lys Leu Glu Asn Pro Lys Met Ser Leu
380 385 390
Glu Asn Asp Lys Ile Leu Gln Ile Ile Thr Glu Leu Ile Lys Thr
395 400 405
Glu Asn Asn

<210> 14
<211> 351
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3093845CD1

<400> 14
Met Ile Pro Lys Ser Tyr Thr Glu Glu Asp Leu Arg Glu Lys Phe
1 5 10 15
Lys Val Tyr Gly Asp Ile Glu Tyr Cys Ser Ile Ile Lys Asn Lys
20 25 30
Val Thr Gly Glu Ser Lys Gly Leu Gly Tyr Val Arg Tyr Leu Lys
35 40 45
Pro Ser Gln Ala Ala Gln Ala Ile Glu Asn Cys Asp Arg Ser Phe
50 55 60
Arg Ala Ile Leu Ala Glu Pro Lys Asn Lys Ala Ser Glu Ser Ser
65 70 75
Glu Gln Asp Tyr Tyr Ser Asn Met Arg Gln Glu Ala Leu Gly His
80 85 90
Glu Pro Arg Val Asn Met Phe Pro Phe Val Gly Glu Gln Gln Ser
95 100 105
Glu Phe Ser Ser Phe Asp Lys Asn Asp Ser Arg Gly Gln Glu Ala
110 115 120
Ile Ser Lys Arg Leu Ser Val Val Ser Arg Val Pro Phe Thr Glu
125 130 135
Glu Gln Leu Phe Ser Ile Phe Asp Ile Val Pro Gly Leu Glu Tyr

```
                140                     145                     150
Cys Glu Val Gln Arg Asp Pro Tyr Ser Asn Tyr Gly His Gly Val
                155                     160                     165
Val Gln Tyr Phe Asn Val Ala Ser Ala Ile Tyr Ala Lys Tyr Lys
                170                     175                     180
Leu His Gly Phe Gln Tyr Pro Pro Gly Asn Arg Ile Gly Val Ser
                185                     190                     195
Phe Ile Asp Asp Gly Ser Asn Ala Thr Asp Leu Leu Arg Lys Met
                200                     205                     210
Ala Thr Gln Met Val Ala Ala Gln Leu Ala Ser Met Val Trp Asn
                215                     220                     225
Asn Pro Ser Gln Gln Gln Phe Met Gln Phe Gly Gly Ser Ser Gly
                230                     235                     240
Ser Gln Leu Pro Gln Ile Gln Thr Asp Val Val Leu Pro Ser Cys
                245                     250                     255
Lys Lys Lys Ala Pro Ala Glu Thr Pro Val Lys Glu Arg Leu Phe
                260                     265                     270
Ile Val Phe Asn Pro His Pro Leu Pro Leu Asp Val Leu Glu Asp
                275                     280                     285
Ile Phe Cys Arg Phe Gly Asn Leu Ile Glu Val Tyr Leu Val Ser
                290                     295                     300
Gly Lys Asn Val Gly Tyr Ala Lys Tyr Ala Asp Arg Ile Ser Ala
                305                     310                     315
Asn Asp Ala Ile Ala Thr Leu His Gly Lys Ile Leu Asn Gly Val
                320                     325                     330
Arg Leu Lys Val Met Leu Ala Asp Ser Pro Arg Glu Glu Ser Asn
                335                     340                     345
Lys Arg Gln Arg Thr Tyr
                350


<210> 15
<211> 472
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3685685CD1

<400> 15
Met Gly Gln Ser Gly Arg Ser Arg His Gln Lys Arg Ala Arg Ala
  1               5                      10                      15
Gln Ala Gln Leu Arg Asn Leu Glu Ala Tyr Ala Ala Asn Pro His
                 20                      25                      30
Ser Phe Val Phe Thr Arg Gly Cys Thr Gly Arg Asn Ile Arg Gln
                 35                      40                      45
Leu Ser Leu Asp Val Arg Arg Val Met Glu Pro Leu Thr Ala Ser
                 50                      55                      60
Arg Leu Gln Val Arg Lys Lys Asn Ser Leu Lys Asp Cys Val Ala
                 65                      70                      75
Val Ala Gly Pro Leu Gly Val Thr His Phe Leu Ile Leu Ser Lys
                 80                      85                      90
Thr Glu Thr Asn Val Tyr Phe Lys Leu Met Arg Leu Pro Gly Gly
                 95                     100                     105
Pro Thr Leu Thr Phe Gln Val Lys Lys Tyr Ser Leu Val Arg Asp
                110                     115                     120
```

```
Val Val Ser Ser Leu Arg Arg His Arg Met His Glu Gln Gln Phe
                125                 130                 135
Ala His Pro Pro Leu Leu Val Leu Asn Ser Phe Gly Pro His Gly
                140                 145                 150
Met His Val Lys Leu Met Ala Thr Met Phe Gln Asn Leu Phe Pro
                155                 160                 165
Ser Ile Asn Val His Lys Val Asn Leu Asn Thr Ile Lys Arg Cys
                170                 175                 180
Leu Leu Ile Asp Tyr Asn Pro Asp Ser Gln Glu Leu Asp Phe Arg
                185                 190                 195
His Tyr Ile Lys Val Val Pro Val Gly Ala Ser Arg Gly Met Lys
                200                 205                 210
Lys Leu Leu Gln Glu Lys Phe Pro Asn Met Ser Arg Leu Gln Asp
                215                 220                 225
Ile Ser Glu Leu Leu Ala Thr Gly Ala Gly Leu Ser Glu Ser Glu
                230                 235                 240
Ala Glu Pro Asp Gly Asp His Asn Ile Thr Glu Leu Pro Gln Ala
                245                 250                 255
Val Ala Gly Arg Gly Asn Met Arg Ala Gln Gln Ser Ala Val Arg
                260                 265                 270
Leu Thr Glu Ile Gly Pro Arg Met Thr Leu Gln Leu Ile Lys Val
                275                 280                 285
Gln Glu Gly Val Gly Glu Gly Lys Val Met Phe His Ser Phe Val
                290                 295                 300
Ser Lys Thr Glu Glu Glu Leu Gln Ala Ile Leu Glu Ala Lys Glu
                305                 310                 315
Lys Lys Leu Arg Leu Lys Ala Gln Arg Gln Ala Gln Gln Ala Gln
                320                 325                 330
Asn Val Gln Arg Lys Gln Glu Gln Arg Glu Ala His Arg Lys Lys
                335                 340                 345
Ser Leu Glu Gly Met Lys Lys Ala Arg Val Gly Gly Ser Asp Glu
                350                 355                 360
Glu Ala Ser Gly Ile Pro Ser Arg Thr Ala Ser Leu Glu Leu Gly
                365                 370                 375
Glu Asp Asp Asp Glu Gln Glu Asp Asp Asp Ile Glu Tyr Phe Cys
                380                 385                 390
Gln Ala Val Gly Glu Ala Pro Ser Glu Asp Leu Phe Pro Glu Ala
                395                 400                 405
Lys Gln Lys Arg Leu Ala Lys Ser Pro Gly Arg Lys Arg Lys Arg
                410                 415                 420
Trp Glu Met Asp Arg Gly Arg Gly Arg Leu Cys Asp Gln Lys Phe
                425                 430                 435
Pro Lys Thr Lys Asp Lys Ser Gln Gly Ala Gln Ala Arg Arg Gly
                440                 445                 450
Pro Arg Gly Ala Ser Arg Asp Gly Gly Arg Gly Arg Gly Arg Gly
                455                 460                 465
Arg Pro Gly Lys Arg Val Ala
                470
```

<210> 16
<211> 616
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3825977CD1

<400> 16
Met Ser Ser Leu Ala Val Arg Asp Pro Ala Met Asp Arg Ser Leu
1 5 10 15
Arg Ser Val Phe Val Gly Asn Ile Pro Tyr Glu Ala Thr Glu Glu
20 25 30
Gln Leu Lys Asp Ile Phe Ser Glu Val Gly Ser Val Val Ser Phe
35 40 45
Arg Leu Val Tyr Asp Arg Glu Thr Gly Lys Pro Lys Gly Tyr Gly
50 55 60
Phe Cys Glu Tyr Gln Asp Gln Glu Thr Ala Leu Ser Ala Met Arg
65 70 75
Asn Leu Asn Gly Arg Glu Phe Ser Gly Arg Ala Leu Arg Val Asp
80 85 90
Asn Ala Ala Ser Glu Lys Asn Lys Glu Glu Leu Lys Ser Leu Gly
95 100 105
Pro Ala Ala Pro Ile Ile Asp Ser Pro Tyr Gly Asp Pro Ile Asp
110 115 120
Pro Glu Asp Ala Pro Glu Ser Ile Thr Arg Ala Val Ala Ser Leu
125 130 135
Pro Pro Glu Gln Met Phe Glu Leu Met Lys Gln Met Lys Leu Cys
140 145 150
Val Gln Asn Ser His Gln Glu Ala Arg Asn Met Leu Leu Gln Asn
155 160 165
Pro Gln Leu Ala Tyr Ala Leu Leu Gln Ala Gln Val Val Met Arg
170 175 180
Ile Met Asp Pro Glu Ile Ala Leu Lys Ile Leu His Arg Lys Ile
185 190 195
His Val Thr Pro Leu Ile Pro Gly Lys Ser Gln Ser Val Ser Val
200 205 210
Ser Gly Pro Gly Pro Gly Pro Gly Pro Gly Leu Cys Pro Gly Pro
215 220 225
Asn Val Leu Leu Asn Gln Gln Asn Pro Pro Ala Pro Gln Pro Gln
230 235 240
His Leu Ala Arg Arg Pro Val Lys Asp Ile Pro Pro Leu Met Gln
245 250 255
Thr Pro Ile Gln Gly Gly Ile Pro Ala Pro Gly Pro Ile Pro Ala
260 265 270
Ala Val Pro Gly Ala Gly Pro Gly Ser Leu Thr Pro Gly Gly Ala
275 280 285
Met Gln Pro Gln Leu Gly Met Pro Gly Val Gly Pro Val Pro Leu
290 295 300
Glu Arg Gly Gln Val Gln Met Ser Asp Pro Arg Ala Pro Ile Pro
305 310 315
Arg Gly Pro Val Thr Pro Gly Gly Leu Pro Pro Arg Gly Leu Leu
320 325 330
Gly Asp Ala Pro Asn Asp Pro Arg Gly Gly Thr Leu Leu Ser Val
335 340 345
Thr Gly Glu Val Glu Pro Arg Gly Tyr Leu Gly Pro Pro His Gln
350 355 360
Gly Pro Pro Met His His Ala Ser Gly His Asp Thr Arg Gly Pro
365 370 375
Ser Ser His Glu Met Arg Gly Gly Pro Leu Gly Asp Pro Arg Leu
380 385 390

Leu Ile Gly Glu Pro Arg Gly Pro Met Ile Asp Gln Arg Gly Leu
395 400 405
Pro Met Asp Gly Arg Gly Gly Arg Asp Ser Arg Ala Met Glu Thr
410 415 420
Arg Ala Met Glu Thr Glu Val Leu Glu Thr Arg Val Met Glu Arg
425 430 435
Arg Gly Met Glu Thr Cys Ala Met Glu Thr Arg Gly Met Glu Ala
440 445 450
Arg Gly Met Asp Ala Arg Gly Leu Glu Met Arg Gly Pro Val Pro
455 460 465
Ser Ser Arg Gly Pro Met Thr Gly Gly Ile Gln Gly Pro Gly Pro
470 475 480
Ile Asn Ile Gly Ala Gly Gly Pro Pro Gln Gly Pro Arg Gln Val
485 490 495
Pro Gly Ile Ser Gly Val Gly Asn Pro Gly Ala Gly Met Gln Gly
500 505 510
Thr Gly Ile Gln Gly Thr Gly Met Gln Gly Ala Gly Ile Gln Gly
515 520 525
Gly Gly Met Gln Gly Ala Gly Ile Gln Gly Val Ser Ile Gln Gly
530 535 540
Gly Gly Ile Gln Gly Gly Gly Ile Gln Gly Ala Ser Lys Gln Gly
545 550 555
Gly Ser Gln Pro Ser Ser Phe Ser Pro Gly Gln Ser Gln Val Thr
560 565 570
Pro Gln Asp Gln Glu Lys Ala Ala Leu Ile Met Gln Val Leu Gln
575 580 585
Leu Thr Ala Asp Gln Ile Ala Met Leu Pro Pro Glu Gln Arg Gln
590 595 600
Ser Ile Leu Ile Leu Lys Glu Gln Ile Gln Lys Ser Thr Gly Ala
605 610 615
Ser

<210> 17
<211> 112
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 4941262CD1

<400> 17
Met Val Lys Gly Arg Thr Gly Gln Arg Val Arg Leu Tyr Val Arg
1 5 10 15
Gly Thr Ile Leu Gly Tyr Lys Arg Ser Lys Ser Asn Gln Tyr Glu
20 25 30
Thr Thr Ser Leu Ile Gln Ile Glu Gly Val Asn Thr Lys Glu Asp
35 40 45
Val Ala Trp Tyr Ala Gly Lys Arg Met Ala Tyr Ile Tyr Lys Ala
50 55 60
Lys Thr Lys Ser Ser Glu Thr Arg Tyr Arg Cys Ile Trp Gly Lys
65 70 75
Val Thr Arg Pro His Gly Asn Ser Gly Val Val Arg Ala Lys Phe
80 85 90
Lys Ser Asn Leu Pro Pro Glu Ser Met Gly Arg Lys Val Arg Val
95 100 105

Phe Met Tyr Pro Ser Ser Ile
110

<210> 18
<211> 1872
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 399781CB1

<400> 18
gccctctagc tgtgtgtgtc tgaggctcgg ccgcctgagc cgcggacggt ttgctgagcc 60
cgttagtgcg cccggccgag acacgccgcc gccatgtccc gctacctgcg tccccccaac 120
acgtctctgt tcgtcaggaa cgtggccgac gacaccaggt ctgaagactt gcggcgtgaa 180
tttggtcgtt atggtcctat agttgatgtg tatgttccac ttgatttcta cactcgccgt 240
ccaagaggat ttgcttatgt tcaatttgag gatgttcgtg atgctgaaga cgctttacat 300
aatttggaca gaaagtggat ttgtggacgg cagattgaaa tacagtttgc ccagggggat 360
cgaaagacac caaatcagat gaaagccaag gaagggagga atgtgtacag ttcttcacgc 420
tatgatgatt atgacagata cagacgttct agaagccgaa gttatgaaag gaggagatca 480
agaagtcggt cttttgatta caactataga agatcgtata gtcctagaaa cagtagaccg 540
actggaagac cacggcgtag agaagccatt ccgacaatga tagaccaaac tgcagctgga 600
atacccagta cagttctgct tactacactt caagaaagat ctgaaagcgg aaaaagaacc 660
aaagaagggc agttcaagcg accaaagggt gggtggaagg tgctgcagta tgaatactgt 720
acgaatattt tgactctggt ctgaaaagat aaaagaatgt tatcgaaaac tacatggaat 780
aattgaagtc ccttcaagtt tgaaagtaag catttagga caaataaaag gaaattcaac 840
tttgtacttg tggaaactaa tccctaaata tgaataggtt tatattgatt catgggtaac 900
aggtccataa taaattattg gaaactagga tgtctgaata tcaaggaaga cagccatagt 960
ctcttacagt gcctctgttg gtctgtctca aactgaattg ggtgggaaaa ggtatggtcc 1020
aatataaaag ttccattttt gccattattg gcaaatcttg cctttgttta ttttggtgcc 1080
agtgttttct gcttaatcat ttgctttgtt ggcatctgtg tttatttact tgtacaccac 1140
atgcagttta catctgtctt aactactcct tcccaggtaa attccaatta tatttgacat 1200
ccagctaaga gggcccatct cttctcacct ctttcctagt cagtatattc agcaaatatt 1260
tattgagccc ttactgtggg caaatcattg tactggataa ttgagaaaaa tagataattc 1320
ccttattcag taaatgtcta ctgagcacaa tctagtgaat cattacagta tggcctcatt 1380
gttttgtttg aggtgtgtta ttcataacaa tattttacac cattcgtatc aatgtaatta 1440
tagaacacaa tatacgatca aggataagta attgtgtggt tatctgccat ttaaaagtat 1500
ccagtatttg atcacattat tataaataat gaaaaaatga tttaatctgt aataaactgg 1560
tttattgtgc agtgactgta atatactaga gttataataa attgtttact ctgcctcacc 1620
aaacacatgc taggatataa cccccaaaat aagtatttaa ctttgcatta ggtataaagg 1680
agactgggtg ctataattag attattttga ggcagacaga gagctgttat cctaactgat 1740
ttagtatgtt ctgtaattga gaaaatgttc accaaattat actttttagt gatttacatg 1800
tacattttat aggggacatg ttctgtgtat agcgaataaa taacttttat agtatcaaaa 1860
aaaaaaaaaa aa 1872

<210> 19
<211> 5897
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1252206CB1

<400> 19
gactcactat agggaatttt gccctcgagg caagaattcg gaacgaggaa cctcttgagg 60
atcgaatctt cactcccgct gtctcagcag tctacagcac ggtaacacaa gtggcaagac 120
agccgggaac ccctacccca tcccсttatt cagcacatga aataaacaag ggcatccaa 180
atcttgcggc aacgcccccg ggacatgcat cgtcccctgg actctctcaa accccttatc 240
cctctggaca gaatgcaggt ccaaccacgc tggtataccc tcaaacccct cagacaatga 300
attcacaacc tcaaacccgt tctccgtttt tccagaggcc tcaaatacag cctcctagag 360
ctaccatccc gaacagcagt ccttccattc gtcctggtgc acagacaccc actgcagtgt 420
accaggctaa tcagcacatc atgatggtta accatctgcc catgccgtac ccagtgcccc 480
aggggcctca gtactgtata ccacagtacc gtcatagtgg ccctccttat gttgggcccc 540
cccaaaaata tccagttcaa ccaccggggc caggtccttt ttatcctgga ccaggacctg 600
gggacttccc caatgcttat ggaacgcctt tttacccaag tcagccggtg tatcagtcag 660
cacctatcat agtgcctacg cagcaacagc cgcctccagc caagagagag aaaaaaacta 720
taagaattcg ggatccaaac cagggaggta aagacataac agaggagatt atgtctggag 780
gtggcagcag aaatcctact ccacccatag gaagacccac gtccacacct actcctcctc 840
agctgcccag ccaggtcccc gagcacagcc ctgtggttta tgggactgtg gagagcgctc 900
atcttgctgc cagcacccct gtcactgcag ctagcgacca gaagcaagag gagaagccaa 960
aaccagatcc agtgttaaag tctccttccc cagtccttag gctagtcctc agtggagaga 1020
agaaagaaca agaaggccag acatctgaaa ctactgcaat agtatccata gcagagcttc 1080
ctctgcctcc atcacctacc actgtttctt ctgttgctcg aagtacaatt gcagccccca 1140
cctcttctgc tcttagtagc caaccaatat tcaccactgc tatagatgac agatgtgaac 1200
tctcatcccc aagagaagac acaattccta tacccagcct cacatcttgc acagaaacat 1260
cagacccttt accaacaaat gaaaatgatg atgatatatg caagaaaccc tgtagtgtag 1320
cacctaatga tattccactg gtttctagta ctaacctaat taatgaaata aatggagtta 1380
gcgaaaaatt atcagccacg gagagcattg tggaaatagt aaaacaggaa gtattgccat 1440
tgactcttga attggagatt ctcgaaaatc cccccagaaga aatgaaactg gagtgtatcc 1500
cagctcccat cacccсttcc acagttcctt cctttcctcc aactcctcca actcctccag 1560
cttctcctcc tcacactcca gtcattgttc ctgctgctgc cactactgtt agttctccga 1620
gtgctgccat cacagtccag agagtcctag aggaggacga gagcataaga acttgcctta 1680
gtgaagatgc aaaagagatt cagaacaaaa tagaggtaga agcagatggg caaacagaag 1740
agattttgga ttctcaaaac ttaaattcaa gaaggagccc tgtcccagct caaatagcta 1800
taactgtacc aaagacatgg aagaaaccaa aagatcggac ccgaaccact gaagagatgt 1860
tagaggcaga attggagctt aaagctgaag aggagctttc cattgacaaa gtacttgaat 1920
ctgaacaaga taaaatgagc caggggtttc atcctgaaag agacccctct gacctaaaaa 1980
aagtgaaagc tgtggaagaa aatggagaag aagctgagcc agtacgtaat ggtgctgaga 2040
gtgtttctga gggtgaagga atagatgcta attcaggctc cacagatagt tctggtgatg 2100
gggttacatt tccatttaaa ccagaatcct ggaagcctac tgatactgaa ggtaagaagc 2160
agtatgacag ggagtttctg ctggacttcc agttcatgcc tgcctgtata caaaaaccag 2220
agggcctgcc tcctatcagt gatgtggttc ttgacaagat caaccaaccc aaattgccaa 2280
tgcgaactct ggatcctcga attttgcctc gaggaccaga ctttacacca gcctttgctg 2340
attttggaag gcagacacct ggtggaagag gcgtaccttt gttgaatgtt gggtcacgaa 2400
gatctcaacc tggccaaaga agagaaccca gaaagatcat cacagtttct gtaaaagaag 2460
atgtacacct gaaaaaggca gaaaatgcct ggaagccaag ccaaaaacga gacagccaag 2520
ccgatgatcc cgaaaacatt aaaacccagg agcttttag aaaagttcga agtatcttaa 2580
ataaattgac accacagatg ttcaatcaac tgatgaagca agtgtcagga cttactgttg 2640
acacagagga gcggctgaaa ggagttattg acctggtctt tgagaaggct attgatgaac 2700
ccagtttctc tgtggcttac gcaaacatgt gtcgatgtct agtaacgctg aaagtaccca 2760
tggcagacaa gcctggtaac acagtgaatt tccggaagct gctactgaac cgttgccaga 2820
aggagtttga aaaagataaa gcagatgatg atgtctttga gaagaagcag aaagaacttg 2880
aggctgccag tgctccagag gagaggacaa ggcttcatga tgaactggaa gaagccaagg 2940
acaaagcccg gcggagatcc attggcaaca tcaagtttat tggagaactc tttaaactca 3000
aaatgctgac tgaagccatc atgcatgact gtgtggtgaa gctgctaaag aaccatgatg 3060
aagaatccct ggagtgcctg tgtcgcctgc tcaccaccat tggcaaagac ttggactttg 3120
aaaaagcaaa gccacgtatg gaccagtact ttaatcagat ggagaaaatt gtgaaagaaa 3180
aaaaaacctc atctaggatt cggttcatgc ttcaagatgt tatagaccta aggctgtgca 3240
attgggtatc tcgaagagca gatcaagggc ctaaaactat cgaacagatt cacaaagagg 3300
ctaaaataga agaacaagaa gagcaaagga aggtccagca actcatgacc aaagagaaga 3360

```
gaagaccagg tgtccagaga gtggacgaag gtgggtggaa cactgtacaa ggggccaaga 3420
acagtcgggt actggacccc tcaaaattcc taaaaatcac taagcctaca attgatgaaa 3480
aaattcagct ggtacctaaa gcacagctag gcagctgggg aaaaggcagc agtggtggag 3540
caaaggcaag tgagactgat gccttacggt caagtgcttc cagtttaaac agattctctg 3600
ccctgcaacc tccagcaccc tcagggtcca cgccatccac gcctgtagag tttgattccc 3660
gaaggacctt aactagtcgt ggaagtatgg gcagggagaa gaatgacaag cccctccat 3720
ctgcaacagc tcggccaaat actttcatga ggggtggcag cagtaaagac ctgctagaca 3780
atcagtctca agaagagcag cggagagaga tgctggagac cgtgaagcag ctcacaggag 3840
gtgtggatgt ggagaggaac agcactgagg ctgagcgaaa taaaacaagg gagtcagcaa 3900
aaccagaaat ttcagcaatg tcagctcatg acaaggctgc attatcagaa gaggaactgg 3960
agaggaagtc gaaatctatc attgatgaat ttctacacat taatgatttt aaggaagcca 4020
tgcagtgtgt ggaagagctg aatgcccagg gcctactaca tgtttttgtg agagtgggag 4080
tggagtccac cctggaaagg agccagatca ccagggatca catgggccaa ctactctatc 4140
agctggtaca gtcagaaaaa ctcagcaaac aggacttttt caaaggtttt tcagaaactt 4200
tggaattggc agatgacatg gccattgata ttccccatat ttggttgtac cttgctgaac 4260
tggtgacccc catgttaaaa gaaggtggaa tctccatgag agaacttacc atagaattta 4320
gcaaaccttt acttcctgtt ggaagagctg gggtcttgct atctgaaata ttgcacctac 4380
tatgcaaaca aatgagccat aagaaagtgg gagccttatg gagggaggct gacctcagct 4440
ggaaggactt tttaccagaa ggagaagatg tacataattt tcttttggag cagaagttgg 4500
acttcataga gtctgacagt ccctgttcct ctgaagcact ttcaaagaaa gaactgtctg 4560
ccgaagagct gtataagcga ctcgagaaac tcattattga ggacaaagcg aatgatgaac 4620
agatctttga ctgggtagag gctaatctag acgaaatcca gatgagttca cctacattcc 4680
ttagagcttt aatgactgct gtttgtaaag cagctattat agccgactct tctaccttca 4740
gagtggacac tgctgttatc aagcagagag tgccgatctt actcaagtac ctagactcag 4800
atacagagaa ggaactgcaa gcactttatg cactacaagc atcgatagta aaacttgatc 4860
aacctgccaa tttgctgcgg atgtttttg attgtctata tgacgaggag gtgatctccg 4920
aggatgcctt ctacaaatgg gagagcagca aggaccctgc agagcagaat gggaagggcg 4980
tggctctgaa atctgtcacg gcattcttca cgtggctgcg ggaagcagaa gaggagtctg 5040
aggataacta aaacttcaaa tacacaaaat gaaacaaaag aaacaattta agtatttttt 5100
taaaaagttt cacgtcttcg ccaatcacag tgcagcaagg ccaattctcg cagaaacccc 5160
cacgtgtgca cgagtgggag aggggaaaga gaaaaaaagg tgatcatgga ggaaaaaggt 5220
actggataaa agtaaacttc aaaccttagg gcgggagcac taaaaccaaa atacatgtat 5280
tatttataga aaatatttc tgttttaatc ttttcttttt aaacaaggac tcatacttaa 5340
aaaaatgttt agcaaaaaaa aaaaaagttg agaactttta atttatttta aggactgcaa 5400
atgccagtgt aatttttaa tttgcagttt ctgtaaacaa cttgtataat agaaaagcag 5460
agaaataaat ttccctcccc ttcaagatgc acctcatgtt tgttttaagg tatagcattt 5520
agtccagatt tgagaaagtt tggggtgaac aaggtaagaa agattttttt ttttttggca 5580
tcaaatcttt ctgcctgcct ctcagcttgc ttcagaaaat ttaaaaaatc acaatagtaa 5640
tcaaaacata cataacattg aaacagaagg aaatgctgtg gaccacagaa ctccaagaat 5700
tgtttaaaaa aaaaaaagtg ctaccctgag aaaagtactc ttaatactct tgaaatcttt 5760
agagcaactt taaggcttgt aaatacatag aacaaatatt taaaaaaaca aaaagaaatt 5820
gactcagtac tatttctttt cactttgaaa atataaagaa caaaataaag acaaacattg 5880
caagtttaaa aaaaaaa                                                5897
```

<210> 20
<211> 919
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2950994CB1

<400> 20

```
tggggtgggg ggccacccca agagctcaga gcctcctgac tcctgtgccc tctgtcccag  60
gtggagatgt ttggggtgac aggccctggg ctggagcaga gctcacagct gctggaggag 120
```

```
ttcctgtccc ttcagatgga gatcttgaca gagctgggct tgcacttccg ggtcctggat 180
atgcccaccc aagaactggg cctccccgcc taccgcaagt ttgacattga ggcctggatg 240
ccaggccgag gccgctttgg agaggtcacc agtgcttcca actgcacaga cttccagagc 300
cgccgcctcc acatcatgtt ccagaccgag gctggggagc tgcagtttgc ccacacggtg 360
aacgccaccg cctgtgctgt cccccgcctt ctcatcgcgc tcctggagag taaccagcag 420
aaggacggct cagtgctcgt gccccctgcc ctccagtcct acctcggcac tgatcggatc 480
acagccccta cccacgtgcc tctccagtac atcggcccca accagccccg gaagcctggg 540
ctgcctggcc agcctgctgt aagctaagaa cccacccaca gcagccctcg ggggtgtcac 600
tgcttcctgg agttcaggag accccggaca cctgggacct gtgttgctga gcccgtcctg 660
acatctgtgt tcttcctgtc agctccacgc ccgggcccct ggaccacggg gtccacctct 720
cctctgtcct tgctgcctca gagtcagtca ctgaccctgt tatcattgag ggtcccagtg 780
ggaagcagga cgtctgggct ttacggttct agggacagga gaagcagagg aagaggcttc 840
catccctcct tccttctttc ctcctacagt gctgagcaaa aagtccccaa taaatggtca 900
ggacaaagaa aaaaaaaaa                                              919
```

<210> 21
<211> 1867
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3461657CB1

<400> 21

```
cggggcacca cagccatgtg ctcgttagcg tcaggcgcta ccggcggccg gggcgctgtg   60
gagaatgagg aggacctgcc agaactgtcg gacagcgggg acgaggccgc ctgggaggat  120
gaggacgatg cagatctccc ccacggcaag cagcagaccc cctgcctgtt ctgtaacagg  180
ttattcacat ctgctgaaga aacattttca cactgtaagt ctgagcatca gtttaatatt  240
gacagcatgg ttcataaaca tggacttgaa ttttatggat acattaagct aataaatttt  300
attagactta agaatcctac agttgagtac atgaattcca tatacaaccc agtgccttgg  360
gagaaagaag agtatttgaa gccagtatta gaagatgacc ttttacttca atttgatgta  420
gaagatcttt atgaaccggt gtcagtaccc ttctcatacc ccaatggact cagtgaaaat  480
acatctgttg ttgaaaaatt gaaacatatg gaagccaggg cactgtctgc tgaagccgca  540
ttggccagag cacgtgagga tctgcaaaaa atgaaacaat ttgctcagga ttttgtgatg  600
cacacagatg tcagaacctg ctcgtcatct actagtgtca ttgcggacct ccaggaggat  660
gaggatggtg tttatttcag ctcatacggg cattatggga tacatgaaga aatgctaaag  720
gacaaaatac gaacagaaag ctaccgagat ttcatatacc aaaatccaca tatcttcaaa  780
gacaaggtag ttttggatgt tgggtgtgga actggaattc tctctatgtt tgctgctaaa  840
gctggggcga agaaggttct tggagttgat caatctgaaa tactttacca ggcaatggat  900
attataagac taaataaact tgaagatact attacactaa ttaaaggaaa gattgaagaa  960
gttcatcttc ctgtagaaaa agtagatgtt atcatatctg agtggatggg ctattttctt 1020
ctgtttgagt ctatgttaga ttctgtcctt tatgcaaaga acaaatactt ggcaaaagga 1080
ggctcggtct accctgacat ttgcactatc agccttgtag cagtgagtga tgtgaataaa 1140
catgctgata gaattgcttt ttgggatgat gtctatggct tcaagatgtc ctgcatgaag 1200
aaagcagtta ttccagaagc tgttgtggaa gttttagatc cgaagactct tatttcagaa 1260
ccttgtggta ttaagcatat agattgccat acgacgtcta tctcagattt ggaattttca 1320
tcagatttta ccctgaaaat cacaaggaca tccatgtgca cggcaattgc tggctacttt 1380
gatatatatt ttgagaagaa ttgccacaac agggtcgtgt tctctacggg ccctcagagc 1440
accaaaacac actggaaaca aacagtattt ctactggaaa aaccattttc agttaaagca 1500
ggtgaagcct tgaaaggaaa ggtcacagtt cacaagaata agaaagatcc acgttctctc 1560
accgtgaccc tcacgttgaa taattcaact caaacttatg gtctccagtg aaacagccat 1620
aaaagcacac taccttgtag tttttaatgt gggggtagag tgggtcagca ggagggagct 1680
ggttttatgt gagcagatgg atggatgatg gaccctttcc taatgagcct cctcaataag 1740
agagaagttc tcattgtggg aatctgacat agttcagctg aggaagagaa tcagctgatc 1800
```

ctcatggtct gccacgtaat cattttctta gacgtttgct ccaccagatt taaccaaatg 1860
taactcc 1867

<210> 22
<211> 702
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 053076CB1

<400> 22
gaagtgagtg atcgaaagca tggcgtcggt ggtgttggcg ctgaggaccc ggacagccgt 60
tacatccttg ctaagcccca ctccggctac agctcttgct gtcagatacg catccaagaa 120
gtcgggtggt agctccaaaa acctcggtgg aaagtcatca ggcagacgcc aaggcattaa 180
gaaaatggaa ggtcactatg ttcatgctgg gaacatcatt gcaacacagc gccatttccg 240
ctggcaccca ggtgcccatg tgggtgttgg gaagaataaa tgtctgtatg ccctggaaga 300
ggggatagtc cgctacacta aggaggtcta cgtgcctcat cccagaaaca cggaggctgt 360
ggatctgatc accaggctgc ccaagggtgc tgtgctctac aagacttttg tccacgtggt 420
tcctgccaag cctgagggca ccttcaaact ggtagctatg ctttgatgtc ctgttgaggc 480
catcggacag agactggagc ccaggtgaca ggagatggtg ataccagaag tcaagggttg 540
gggtggcgac acggcctccc gaggaagagg tctgcttgat ggtgactctg caggagactc 600
tgaagtgact gctgggaaac cctttgggag acctgacctg gggccaaaaa taaagtgagc 660
cagcgtcatg aacgcatgct atttagggac aaaaaaaaaa aa 702

<210> 23
<211> 2459
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1292379CB1

<400> 23
tgcaaaagaa acaaagcaca ttggctagta gcaataaaaa aagtaaagtc aatacagcat 60
tgaggaattt aaggtatcgt cggggcattc acaaatgcat tgagtgttgt tccgaaataa 120
aagattttgc aaaccacttt cctacgtacg tccactgtag tttttgcaga tacaacacta 180
gctgtagcaa agcctatgta aatcatatga tgagctttca tagtaaccgt ccaagcaaaa 240
ggttttgtat ttttaagaag cattcagaaa atctccgggg cattactcta gtgtgcctta 300
attgtgattt cctaagtgat gtttctggct tagataatat ggctacacac ttaagtcaac 360
ataaaactca tacttgccaa gttgtaatgc agaaagtttc tgtttgtatc ccaacttctg 420
agcacctttc tgaattaaaa aaagaagctc ccgcaaagga acaagaacct gtgtctaagg 480
aaattgcaag acctaacatg gctgaaagag aaacagaaac atcaaattct gaaagtaaac 540
aagataaagc tgcttcttca aaagaaaaaa atggatgtaa tgcaaattca tttgaaggct 600
catcaacaac aaaaagtgaa gaaagcataa cagtttcaga taaggaaaat gaaacctgtc 660
ttgcagacca ggaaactggc tcaaaaaaca tcgtcagttg tgattcaaat attggtgcag 720
ataaagtgga aaagaaaaaa caaatacaac acgtttgtca ggaaatggag ttgaagatgt 780
gccaaagttc agaaaacata atcttatctg atcagattaa agatcacaac tccagtgaag 840
ccagattttc ttcaaagaat attaaggatt tgcgattagc atcagataat gtaagcattg 900
atcagttttt gagaaaaaga catgaacctg aatctgttag ttctgatgtt agcgagcaag 960
gcagtattca tttggaacct ctgactccat ccgaggtact tgagtatgaa gccacagaga 1020
ttcttcagaa aggtagtggt gatccttcag ccaagactga tgaagtagtg tctgatcaaa 1080
cagatgacat tcctggagga aataacccta gcacaacaga ggcaacagta gacctggaag 1140

atgaaaaaga aagaagttga aattagtcat tttaagtttc agtgtaccaa cgataagggc 1200
atttggaaca gtgctatcag gtgagctcag tggtgctgtt gtaggttcag aaatggaaat 1260
atgtaaggga ggtcacacat acactttacc tgtatgttca acctatgtta tcaaacaaac 1320
caattcacca ataatagcat gattagtagg gattcccaaa aagtttttaa aaacacgaac 1380
aggattttaa tgataattaa atttgcagtg gaaaggtctc atttaatggt tttcaaggaa 1440
atgggatttg gttgctgaca tgaattgatg atattagtaa tatttataaa gcctttcaaa 1500
cttccatcaa tcctaagcta aaaatcttta ttacctgtat atccttttca gttaactgag 1560
aggaagggat ttggaaacca tgtacttttg gggagtaatt gattaaaaac aatggctgat 1620
tggcattgtt aatgaaggct ttatttgtga ggatgatgct ggtaaatgga gcatgcttag 1680
agtactaaat tgatctaatg agaatttgga tgaacataaa cttaatttg gatttaatat 1740
aacattccag tcagacgcat gtaaacagaa tatttgaatc tttgtacctc catacaagtg 1800
ttagcctgcc aggctgtaag cttaccttaa ttaaactttc agtgaaagtg gaattattaa 1860
gatataaatt tatatttgtg ctttttgtca gtgtgtaagc tgtgtagaaa ttctttgatg 1920
tattagttgt attaatgtaa agtagaaacc cattgttgaa actcctgtag ctattatgct 1980
tttaatattg ttttaatgat cttccttaga aataggccca taaaaatggt ctggaagcca 2040
aaccaaagta tggtataatg tagatattgt aaagcagtaa actgaaaaca tgtcctggca 2100
tgtattcagc catgtttaag tgacttttct gtaattgtaa aataaaaact tcaaatggga 2160
cctaaaacag tgatgtaaaa gaactggttt tggaaattta gcctaattta tctataagat 2220
ggctgctaaa ttgatttttc agttcttttt atcatctaga atataataga tatagaaatg 2280
aataatatga agaacagtag tttgctttga aatactaata aacttttatt taaaatgctt 2340
catttttact tcttaaaatg tgctttggat tcttaaattt tgtttcactg aatgttcaat 2400
gttttaaatg gcgattaaaa tactctgctg tatatagtag tttttgagta aatatttcc 2459

<210> 24
<211> 1015
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1437783CB1

<400> 24
gaaggggact cgccgcagat ccgctgtact tgcgtccgct acagtatgtc aatcgcttgc 60
cccagcacag ctgtcattct tctctacaga agagcttctc ctcatcaact ggggatgatt 120
acagttcttc ctaaaaaagg atggctgctc tttttctaaa gaggttaaca ctacaaactg 180
taaagtctga aaatagttgc attagatgtt ttggtaaaca catcctgcaa aagacagcac 240
cagcacagtt gtcccctatt gcttctgccc caagactctc cttcctaatt catgcaaaag 300
cctttagtac cgctgaagac acccagaatg aaggaaaaaa gacaaaaaag aataaaacag 360
cttttagtaa cgttggaaga aaaattagtc agcgagttat tcacttattt gatgagaagg 420
gcaatgattt gggaaacatg caccgagcaa atgtgattag acttatggat gagcgagacc 480
tgcgactggt tcaaaggaac accagcacag aacctgcaga gtatcagctc atgacaggat 540
tgcagatcct ccaggagcgg cagaggctga gggagatgga gaaggcgaac cccaaaactg 600
gaccaaccct gagaaaggaa ctgattttgt cttcaaatat tggacaacat gatttggaca 660
caaagactaa acagattcag cagtggatta agaaaaaaca cctagtccag attaccataa 720
agaaaggaaa aaatgtagac gtgtcagaaa atgaaatgga ggagatattt catcaaatac 780
tccagactat gcctggaata gctacattct catctaggcc acaagctgtt caaggaggaa 840
aagctttaat gtgtgttctt cgtgctttga gcaaaaatga ggagaaggca tataaagaaa 900
ctcaagagac ccaggaaaga gacactttga acaaagacca tggaaatgat aaggaatcaa 960
atgttctgca tcagtaattt taataaagaa aagcatgctc tgagaaaaaa aaaaa 1015

<210> 25
<211> 2211
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1557635CB1

<400> 25
```
taaacattac gagattggct tggattctgt cggatggact tggggctagc tgcggcgggg   60
ctggaggagg ccagataacc atgtcagcca cagttgtaga tgcagttaat gctgcacccc  120
tatcggggtc caaagaaatg agtttggaag aaccaaagaa gatgaccaga gaggactgga  180
gaaagaagaa ggagctagaa gaacagcgaa aattgggcaa tgctcctgca gaagttgatg  240
aagaaggaaa agacatcaac ccccatattc ctcagtatat ttcttcagtg ccatggtata  300
ttgatccttc aaaaagacct actttaaaac accagagacc acaaccagaa aaacaaaagc  360
agttcagctc atctggagaa tggtacaaga ggggtgtaaa agagaattcc ataattacta  420
agtaccgcaa aggagcatgt gaaaattgtg ggccatgac acacaaaaag aaagactgct  480
ttgagagacc taggcgagtt ggagccaaat ttacaggtac taatatagct ccagatgaac  540
atgtccagcc tcaactgatg tttgactatg atgggaagag ggatcggtgg aatggctaca  600
atccagaaga acacatgaaa attgttgaag agtatgccaa agttgatttg gcaaaacgaa  660
cattgaaagc ccagaaactc caagaggaat tagcctcagg aaaattagtg gaacaggcta  720
attctccaaa acaccagtgg ggagaagagg aaccaaattc tcagacggaa aaagatcata  780
atagtgaaga tgaggatgaa gataaatatg cagatgatat tgacatgcct ggacagaatt  840
ttgactccaa gagacgaatt actgtccgga atctcaggat tcgagaagat attgcaaaat  900
atttgaggaa tttagatcca aattctgcct actatgatcc aaaaactaga gcaatgagag  960
agaatcctta tgccaatgca ggaaagaatc cagatgaagt gagttatgct ggagataact 1020
ttgttaggta cacaggagat accatttcaa tggctcagac acagttgttt gcatgggaag 1080
cctatgacaa gggatctgaa gtgcatctac aggcagatcc tacaaagcta gagctgttgt 1140
ataagtcctt caaagtcaaa aaagaagatt tcaaagaaca gcagaaagaa agcatcctgg 1200
aaaagtatgg tggccaagaa catttggatg cccctccagc tgaattgctt ttagcccaga 1260
ctgaagacta tgtggagtac tcaagacatg ggacagtcat caaaggacag gagcgggctg 1320
ttgcctgctc taagtatgag gaggatgtga agatccacaa tcacacacat atctggggat 1380
cgtactggaa agaaggccga tggggataca aatgctgtca ctcttttttc aagtattcct 1440
attgtactgg agaagctggg aaggagattg ttaactctga ggagtgtatt ataaatgaga 1500
taactgggga agaatctgtg aaaaaacctc aaaccctcat ggagctgcat caagaaaaac 1560
tgaaagagga aaagaagaag aagaaaaaga aaaagaagaa gcatcgaaag agcagttcag 1620
atagtgatga tgaagaaaag aagcatgaaa aattgaaaaa ggcactgaac gcagaggagg 1680
cccgccttct tcatgtcaag gagaccatgc agattgatga gaggaagcgg ccttacaata 1740
gcatgtatga aactcgagaa cctactgaag aggaaatgga ggcatataga atgaaacgtc 1800
agaggccaga tgaccccatg gcctctttcc ttggacagta gcaactagtc agaagaccat 1860
ccaagataga tgcagctgat acattctttt cagcttctta ttgatgattg tagatagaaa 1920
aatccttgtt tattcttctt gctgcctggc tttaataaat atttcagatg cctcacagta 1980
agttcactcc tttccatact gaggaaacaa gaaaagaaga agaggcacat gaagtgtgct 2040
tttgggaata gaatttaaaa ttggattaag attttatttc cagttttttt tatttattta 2100
tttttttttt tgagacggag tcttgctctg tcgcccaggc tgaagtgcgg tggcgcgatc 2160
tcggctcact gcaagctcca cctcccaggt tcacgccatt ctcctgcctc a          2211
```

<210> 26
<211> 1446
<212> DNA
<213> Homo sapiens

<220>
<221>
<222> 1437
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte ID No.: 2049352CB1

```
<400> 26
ttgccttggg tcctgccgca cagagcggcc tgtctttatc agaggtccct ctgccagggg   60
gagggcccca gagaaaacca gaaagagggt gagagactga ggaagataaa gcgtcccagg  120
gcctcctaca ccagcgcctg agcaggaagc gggaggggcc atgactacga ggccctggga  180
ggtcacttta gggagggctg tcctaaaacc agaagcttgg agcagaaagt gaaaccctgg  240
tgctccagac aaagatctta gtcgggacta gccggccaag gatgaagcct cacttcagaa  300
acacagtgga gcgaatgtat cgagacacat tctcctacaa cttttataat agacccatcc  360
tttctcgtcg gaataccgtc tggctgtgct acgaagtgaa aacaaagggt ccctcaaggc  420
cccctttgga cgcaaagatc tttcgaggcc aggtgtattc cgaacttaag taccacccag  480
agatgagatt cttccactgg ttcagcaagt ggaggaagct gcatcgtgac caggagtatg  540
aggtcacctg gtacatatcc tggagcccct gcacaaagtg tacaagggat atggccacgt  600
tcctggccga ggacccgaag gttaccctga ccatctttgt tgcccgcctc tactacttct  660
gggacccaga ttaccaggag gcgcttcgca gcctgtgtca gaaaagagac ggtccgcgtg  720
ccaccatgaa gatcatgaat tatgacgaat ttcagcactg ttggagcaag ttcgtgtaca  780
gccaaagaga gctatttgag ccttggaata atctgcctaa atattatata ttactgcaca  840
tcatgctggg ggagattctc agacactcga tggatccacc cacattcact ttcaacttta  900
acaatgaacc ttgggtcaga ggacggcatg agacttacct gtgttatgag gtggagcgca  960
tgcacaatga cacctgggtc ctgctgaacc agcgcagggg ctttctatgc aaccaggctc 1020
cacataaaca cggtttcctt gaaggccgcc atgcagagct gtgcttcctg gacgtgattc 1080
ccttttggaa gctggacctg gaccaggact acagggttac ctgcttcacc tcctggagcc 1140
cctgcttcag ctgtgcccag gaaatggcta aattcatttc aaaaaacaaa cacgtgagcc 1200
tgtgcatctt cactgcccgc atctatgatg atcaaggaag atgtcaggag ggctgcgca  1260
ccctggccga ggctggggcc aaaatttcaa tactgacata cagtgaattt aagcactgct 1320
gggacacctt tgtggaccac cagggatgtc ccttccagcc ctgggatgga ctagaggagc 1380
acagccaagc cctgagtggg aggctgcggg gcattctgca gaatcaggga agctgangga 1440
tgggcc                                                            1446

<210> 27
<211> 1349
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2231663CB1

<400> 27
cacggaggac taacagtaac accgccacgc cggcagcaaa gctcattttg gtccccgccc   60
cgttcctctt tctcttttta actccttccc tctttgcgga ttctagaacg gaaccttttt  120
ttaattcttc ccagtagaaa cgtaggaaca atttcgtgaa cgcaatccgg agtgcccaac  180
atggcggcgg ccgtaaggtg catgggtaga gccttgatac atcatcaaag gcatagcctt  240
tccaagatgg tttatcagac atcactttgt tcttgttctg taaacatccg agtgcccaac  300
agacattttg ctgctgctac aaagtctgca aagaaaacaa aaaaaggtgc taaagaaaaa  360
acaccagatg agaaaaaaga tgaaatagaa aaaataaaag catatcccta tatggaaggc  420
gaacctgagg atgatgtcta tttaaaacgc ttatacccga gacagatata tgaggtggag  480
aaagctgttc acttacttaa gaaatttcaa attcttgact ttactagtcc aaagcaaagt  540
gtttatcttg atttgacact ggatatggca ctgggaaaga agaaaaacgt ggagccattt  600
accagtgttc ttagtttgcc atacccattt gcttccgaaa tcaataaagt tgctgtattt  660
acagagaatg catcagaggt caaaatagcg gaagaaaatg gagctgcatt tgcaggaggc  720
actagtctga tacagaagat ttgggatgat gaaattgttg cagactttta cgtagctgtt  780
ccagaaataa tgcctgaact taatcgatta aggaagaaac tgaataaaaa atatccaaag  840
ctttctcgaa attccattgg ccgtgacatc cccaaaatgc ttgaattatt taaaaatgga  900
catgaaatta aggtagatga agaaagggag aactttctcc agaccaaaat agcaacattg  960
gatatgtcaa gtgaccagat agctgccaat ctgcaagcag ttattaatga agtttgtagg 1020
cacagaccgc tgaatttggg tcccttttgtg gtacgtgctt tccttcgtag ttcaacaagt 1080
gaaggtttat tactgaagat tgatccattg ttgcctaaag aagtaaaaaa tgaagaaagt 1140
```

gaaaaagaag atgcctaaat gtggtgaatt gtgaaattac tttcagtggt ttaagaagca 1200
atggagaaaa tgataataca gcataatttt tacatttgat gtcttgttat tgatcatact 1260
tgaaagtgaa ctttaacatt gaaaaatcgt acagtcattt caagaataag aaaataaaat 1320
tttctctttg tctgaaaaaa aaaaaaaaa 1349

<210> 28
<211> 2596
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2604449CB1

<400> 28
gtggcctgag gagctcagtt ccctcagcgc ccgtagcttc ggcggagtct gcgcgatggg 60
cgacccggaa aggccggaag cggccgggct ggatcaggat gagagatcat cttcagacac 120
caacgaaagt gaaataaagt caaatgaaga gccactccta agaaagagtt ctcgccggtt 180
tgtcatcttt ccaatccagt accctgatat ttggaaaatg tataaacagg cacaggcttc 240
cttctggaca gcagaagagg tcgacttatc aaaggatctc cctcactgga acaagcttaa 300
agcagatgag aagtacttca tctctcacat cttagccttt tttgcagcca gtgatggaat 360
tgtaaatgaa aatttggtgg agcgctttag tcaggaggtg caggttccag aggctcgctg 420
tttctatggc tttcaaattc tcatcgagaa tgttcactca gagatgtaca gtttgctgat 480
agacacttac atcagagatc ccaagaaaag ggaattttta tttaatgcaa ttgaaaccat 540
gccctatgtt aagaaaaaag cagattgggc cttgcgatgg atagcagata gaaaatctac 600
ttttggggaa agagtggtgg cctttgctgc tgtagaagga gtttcttct caggatcttt 660
tgctgctata ttctggctaa agaagagagg tcttatgcca ggactcactt tttccaatga 720
actcatcagc agagatgaag gacttcactg tgactttgct tgcctgatgt tccaatactt 780
agtaaataag ccttcagaag aaagggtcag ggagatcatt gttgatgctg tcaaaattga 840
gcaggagttt ttaacagaag ccttgccagt tggcctcatt ggaatgaatt gcattttgat 900
gaaacagtac attgagtttg tagctgacag attacttgtg gaacttggat tctcaaaggt 960
tttcaggcca gaaaatcctt ttgattttat ggaaaacatt tctttagaag gaaaaacaaa 1020
tttctttgag aaacgagttt cagagtatca gcgttttgca gttatggcag aaaccacaga 1080
taacgtcttc accttggatg cagattttta aaaaacctct cgttttaaaa ctctataaac 1140
ttgtcattgg taatgacaga tgccttaatg tgaagcttat ttataatagc aataaaccta 1200
actggatttg gatgaagaag tcttaatact gacatactgg atttttaatg cactggtttg 1260
ttatttggta ttctatctct ttttccaggc ctccaggttg cacatttatt tattatgttc 1320
aatactttgg ttcttagttc ttaaagaatc aagaagttgt gtaatctttt aaaaatatta 1380
tcttgcagat aaagaaaaaa attaagagtg tgtttacaac tgttttctct tttttacagt 1440
acatgtattt aaatcattgc tataataaag ttaagttcat taggaatata aaaacttgca 1500
gttctatgat agattgcatt tattaaaaat gtttcattgt atcacataga aatatggcca 1560
ggaaggactt gagaagacag tttgatccat tgcttttaga caggactggg ttttgctgtc 1620
caattatata caataatagt ttttcttaca actaagctgg ccccagcctt gtcttgatat 1680
taatacatga aatttttata attgtctcat tgtctcattt agaaacatcc atatttttct 1740
gctttttcta ttgccatttt ttatttgtgc atgaattgat tattgagaaa atgtagcagt 1800
ttgcatattt aaaaattaat cattttgcat tttacattta aatatgctaa catcactgtc 1860
atagaattcc caaatttcat ttgtagatac tgaactaagg gctaatgtca ggagctgatt 1920
tttaatgata aagctgcaga tgggctaaat aaaagccaaa ttaatcctac aatcaggtat 1980
tatgttttta aaccaagttg agtgaattgg tagtggactt gggaaatctt ccccagcaga 2040
atctggatga atggcacaga attgaaatct ctttgtttcc caccatttcc ctttaagtgc 2100
tctgctcctt tgtaaaaagt taaagatttg aaagagaatc tcatattccc gaggcattag 2160
gaagaaagga tttaatccct tcaatttggg gcttaatctt gtttaaaaaa atgtaagtga 2220
agatggaagg ctggagagaa tgattgcttt ttgtacagtt aaataaggtc acaatattct 2280
tacatacttt gttttacaac tgtgttttca ttttttcaaa tgtctggcca tttagcaaag 2340
ttatttacta tttactgtgt acatagaaag ctttattatg tgtggtgtat ctaaattttt 2400
tttgctgaaa tacattatgg tcaatcaagc caagcctgca tgtacagaat ttgttttttt 2460

```
ttcaaataaa ttagttgttt tcttattttt ttggcttagt atgttgaaat aaactatggt 2520
atcttcatca ttttgtacat ttcctttttg aggaaggttt ctttataagt gcaagggtac 2580
cctaataaag gaattt                                                 2596
```

```
<210> 29
<211> 2882
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2604993CB1

<400> 29
ccctcgagcg gctcgagggc tttttgcata gacgcccggg caactgaata caaaaagggc   60
aggcctcctg cgccctcctt cccacccccc ttcctgccct gggcgtggag cgccgaccag 120
gtgtgggctt gggtggttgg ttaccgcctt ttgcactagc agtagcaagg aagggggtg  180
ggcgctcttt ctttttctct tagaagaggg tttagcacag gtttttcgt  tctcacttcc 240
acaccacctt accgcctccc gacccccct  ctccccctcc ccacctatcg tcatgacggc 300
ctctccggat tacttggtgg tgctttttgg gatcactgct ggggccaccg gggccaagct 360
aggctcggat gagaaggagt tgatcctgct gttctggaaa gtcgtggatc tggccaacaa 420
gaaggtggga cagttgcacg aagtgctagt tagaccggat cagttggaac tgacggagga 480
ctgcaaagaa gaaactaaaa tagacgtcga aagcctgtcc tcggcgtcgc agctggacca 540
agccctccga cagtttaacc agtcagtgag caatgaactg aatattggag tagggacttc 600
cttctgtctc tgtactgatg ggcagcttca tgtcaggcaa atcctgcatc ctgaggcttc 660
caagaagaat gtactattac ctgaatgctt ctattccttt tttgatcttc gaaaagaatt 720
caagaaatgt tgccctggtt cacctgatat tgacaaactg gacgttgcca caatgacaga 780
gtatttaaat tttgagaaga gtagttcagt ctctcgatat ggagcctctc aagttgaaga 840
tatggggaat ataattttag caatgatttc agagccttat aatcacaggt tttcagatcc 900
agagagagtg aattacaagt ttgaaagtgg aacttgcagc aagatggaac ttattgatga 960
taacaccgta gtcagggcac gaggtttacc atggcagtct tcagatcaag atattgcaag 1020
attcttcaaa ggactcaata ttgccaaggg aggtgcagca ctttgtctga atgctcaggg 1080
tcgaaggaac ggagaagctc tggttaggtt tgtaagtgag gagcaccgag acctagcact 1140
acagaggcac aaacatcaca tggggacccg gtatattgag gtttacaaag caacaggtga 1200
agatttcctt aaaattgctg gtggtacttc caatgaggta gcccagtttc tctccaagga 1260
aaatcaagtc attgttcgca tgcgggggct cccctttcacg gccacagctg aagaagtggt 1320
ggccttcttt ggacagcatt gccctattac tgggggaaag gaaggcatcc tctttgtcac 1380
ctacccagat ggtaggccaa caggggacgc ttttgtcctc tttgcctgtg aggaatatgc 1440
acagaatgcg ttgaggaagc ataaagactt gttgggtaaa agatacattg aactcttcag 1500
gagcacagca gctgaagttc agcaggtgct gaatcgattc tcctcggccc ctctcattcc 1560
acttccaacc cctcccatta ttccagtact acctcagcaa tttgtgcccc ctacaaatgt 1620
tagagactgt atacgccttc gaggtcttcc ctatgcagcc acaattgagg acatcctgga 1680
tttcctgggg gagttcgcca cagatattcg tactcatggg gttcacatgg ttttgaatca 1740
ccagggccgc ccatcaggag atgcctttat ccagatgaag tctgcggaca gagcatttat 1800
ggctgcacag aagtgtcata aaaaaaacat gaaggacaga tatgttgaag tctttcagtg 1860
ttcagctgag gagatgaact ttgtgttaat gggggcact  ttaaatcgaa atggcttatc 1920
cccaccgcca tgtaagttac catgcctgtc tcctccctcc tacacatttc cagctcctgc 1980
tgcagttatt cctacagaag ctgccattta ccagccctct gtgattttga atccacgagc 2040
actgcagccc tccacagcgt actacccagc aggcactcag ctcttcatga actacacagc 2100
gtactatccc agccccccag gttcgcctaa tagtcttggc tacttcccta cagctgctaa 2160
tcttagcggt gtccctccac agcctggcac ggtggtcaga atgcagggcc tggcctacaa 2220
tactggagtt aaggaaattc ttaacttctt ccaaggttac cagtatgcaa ccgaggatgg 2280
acttatacac acaaatgacc aggccaggac tctacccaaa gaatgggttt gtatttaagg 2340
gccccagcag ttagaacatc ctcagaaaag aagtgtttga aagatgtatg gtgatcttga 2400
aacctccaga cacaagaaaa cttctagcaa attcagggga agtttgtcta cactcaggct 2460
gcagtatttt cagcaaactt gattggacaa acgggcctgt gccttatctt ttggtggagt 2520
```

gaaaaaattt gagctagtga agccaaatcg taacttacag caagcagcat gcagcatacc 2580
tggctctttg ctgattgcaa ataggcattt aaaatgtgaa tttggaatca gatgtctcca 2640
ttacttccag ttaaagtggc atcataggtg tttcctaagt tttaagtctt ggataaaaac 2700
tccaccagtg tctaccatct ccaccatgaa ctctgttaag gaagcttcat ttttgtatat 2760
tcccgctctt ttctcttcat ttccctgtct tctgcataat catgccttct tgctaagtaa 2820
ttcaagcata agatcttgga ataataaaat cacaatctta ggagaaagaa taaaatttta 2880
tt 2882

<210> 30
<211> 1777
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2879070CB1

<400> 30
gaaatattga aggagcagta ttgatctttt taccaggact tgctcatatt cagcagttgt 60
atgatcttct atcaaatgat agaagatttt attctgaacg atataaagtg atagctctgc 120
attctattct ttcaacccaa gatcaagctg cagcattcac acttccccct ccaggagtca 180
ggaagattgt tttagcaacc aatattgcag agacgggtat cactattcct gatgttgtat 240
ttgtaattga tactggaaga acaaaagaaa ataagtacca tgaaagcagt cagatgagtt 300
ctttggtgga gacgtttgtc agtaaagcca gtgctttgca gcgccaggga agagctgggc 360
gggtcagaga tggcttctgt ttccgaatgt acacaagaga aagatttgaa ggctttatgg 420
attattctgt tcctgaaatc ttacgtgtac ctttggagga attatgcctt catattatga 480
aatgtaatct tggttctcct gaagatttcc tctccaaagc cttagatcct cctcagctcc 540
aagtgatcag caatgcaatg aatttgctcc gaaaaattgg agcttgtgaa ttaaatgagc 600
ctaaactgac tccgttgggc caacaccttg cagctttacc tgtgaatgtc aagattggca 660
agatgcttat ttttggtgcc atatttggct gccttgaccc agtggcaaca ctagctgcag 720
ttatgacaga gaagtctcct tttaccacac caattggtcg aaaagatgaa gcagatcttg 780
caaaatcagc tttggccatg gcggattcag accacctgac gatctacaat gcatatctag 840
gatggaagaa agcacgacaa gaaggaggtt atcgttctga aatcacatac tgccggagga 900
actttcttaa tagaacatca ctgttaaccc tagaggatgt aaagcaggag ttaataaagt 960
tggttaaggc agcaggattt tcatcttcca caacttctac cagctgggaa ggaaacagag 1020
cctcacagac cctctcattc caagaaattg cccttcttaa agctgtactg gtggctggac 1080
tgtatgacaa tgtggggaag ataatctata caaagtcagt ggatgttaca gaaaaattgg 1140
cttgcattgt ggagacggcc caaggcaaag cacaagtaca cccatcctca gtaaatcgag 1200
atttgcaaac tcatggatgg ctcttatacc aggagaagat aaggtatgcc agagtgtatt 1260
tgagagaaac taccctaata acccctttc cagttttact ttttggtggt gatatagaag 1320
ttcagcaccg agaacgtctt ctttctattg atggctggat ctattttcag gcccctgtaa 1380
agatagctgt cattttcaag cagctgagag ttctcattga ttcagtttta agaaaaaagc 1440
ttgaaaatcc aaagatgtcc cttgaaaatg acaagattct gcagatcatt acggaattga 1500
taaaaacaga gaataactga aactgaaatt catggtcaac tgctttaaaa attaagatga 1560
agatacagtc atgaaattat ctgaaaatgg gtcatcacat taagtatttc attacttaaa 1620
atgttggtac tagccattaa cttaaaggtg gtgggaaaaa agcacatact ttaaacatgt 1680
ataattttct agttcctttt taatgatgat tattctgaat gtatttgcca ctacatttac 1740
aataaattct ttggtattat gcaaaaaaaa aaaaaaa 1777

<210> 31
<211> 1382
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3093845CB1

<400> 31
tggaagtcac cgagatgttg aagatgaaga acttacaaga atctttgtta tgataccaaa 60
gtcctacaca gaagaagatc tgcgggaaaa atttaaggtg tatggagata tcgagtattg 120
cagcattatt aagaataaag tgactggaga aagtaaaggt ttgggctacg tacgatactt 180
aaaaccatca caagctgccc aagcaataga aaactgtgat cgaagtttta gagcaatctt 240
ggctgaacct aaaaataaag catctgaatc ctctgaacaa gattattata gtaatatgag 300
gcaagaagct ttgggacatg aacctagagt aaatatgttt ccatttgtcg gagaacaaca 360
atctgaattt tcaagttttg acaagaatga tagccgaggc caggaagcaa tctccaaacg 420
cttgtcagtt gtatcaagag ttcctttcac tgaagaacag cttttcagca tttttgatat 480
agtaccagga ttggaatatt gtgaagttca acgagatcct tattcaaatt atggtcatgg 540
agtggttcag tattttaatg tagcatcagc tatttatgca aaatacaaat tacatggatt 600
tcagtaccct cctgggaacc gaataggtgt ttccttcatt gatgatggaa gtaatgcaac 660
agatctcctt agaaaaatgg caacacagat ggtagctgca cagcttgcat caatggtgtg 720
gaataaccca agtcagcaac aatttatgca atttggagga agctctggat cacagttgcc 780
tcaaatccag acagatgttg tacttccatc atgcaaaaaa aaagctcctg ctgaaactcc 840
tgtgaaagaa agacttttta ttgtgtttaa tcctcatcct ttacctttag acgtattaga 900
agatatattc tgtcgttttg gtaacctgat cgaagtttac cttgtgtcag gaaaaaatgt 960
ggggtatgcc aagtatgccg atagaataag tgctaatgat gccattgcca ctctacatgg 1020
aaagattctg aatggggtga gacttaaagt tatgctggca gattcgccaa gagaagaatc 1080
taacaaacgg caaagaactt actgattctt gagaacaaag actaaataat gacataatcc 1140
tcagctgact gactgaaaat gtgactggac gcattccctg tggacagttg acagcttttt 1200
tttttttcca tatacctgat agtctgtgta cagcattgtt ttgtctggga agcagggatt 1260
gctgacatgt atttttgaat ccatacatta atgctaaaac gaatatagta gttgttcctt 1320
agagcaatat gttgttacgt gtagcagaaa taaagttttc tttgcttaaa aaaaaaaaaa 1380
aa 1382

<210> 32
<211> 1828
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3685685CB1

<400> 32
cacgccgtgc cggctctcag gcgccggaag tgagctgcgc agcgccggaa gcggcggacg 60
caggaggcct cgtggaggac acagcagcat gggacagtca gggaggtccc ggcaccagaa 120
gcgcgcccgc gcccaggcgc agctccgcaa cctcgaggcc tatgccgcga acccgcactc 180
gttcgtgttc acgcgaggct gcacgggtcg caacatccgg cagctcagcc tggacgtgcg 240
gcgggtcatg gagccgctca ctgccagccg tctgcaggtt cgtaagaaga actcgctgaa 300
ggactgcgtg gcagtggctg ggcccctcgg ggtcacacac tttctgatcc tgagcaaaac 360
agagaccaat gtctacttta agctgatgcg cctcccagga ggccccacct tgaccttcca 420
ggtgaagaag tactcgctgg tgcgtgatgt ggtctcctca ctgcgccggc accgcatgca 480
cgagcagcag tttgcccacc cacccctcct ggtactcaac agctttggcc cccatggtat 540
gcatgtgaag ctcatggcca ccatgttcca gaacctgttc ccctccatca acgtgcacaa 600
ggtgaacctg aacaccatca agcgctgcct cctcatcgac tacaatcccg actcccagga 660
gctggacttc cgccactaca tcaaagttgt tcctgtgggc gcgagtcgcg ggatgaagaa 720
gctgctccag gagaagttcc ccaacatgag ccgcctgcag gacatcagcg agctgctggc 780
cacgggcgcg gggctgtcgg agagcgaggc agagcctgac ggcgaccaca acatcacaga 840
gctgcctcag gctgtcgctg gccgtggcaa catgcgggcc cagcagagtg cagtgcggct 900
caccgagatc ggcccgcgga tgacactgca gctcatcaag gtccaggagg gcgtcgggga 960

gggcaaagtg atgttccaca gttttgtgag caagacggag gaggagctgc aggccatcct 1020
ggaagccaag gagaagaagc tgcggctgaa ggcgcagagg caggcccagc aggcccagaa 1080
tgtgcagcgc aagcaggagc agcgggaggc ccacagaaag aagagcctgg agggcatgaa 1140
gaaggcacgg gtcgggggta gtgatgaaga ggcctctggg atcccttcaa ggacggcgag 1200
cctggagttg ggtgaggacg atgatgaaca ggaagatgat gacatcgagt atttctgcca 1260
ggcggtgggc gaggcgccca gtgaggacct gttccccgag gccaagcaga aacggcttgc 1320
caagtctcca gggcggaagc ggaagcggtg ggaaatggat cgaggcaggg gtcgcctttg 1380
tgaccagaag tttcccaaga ccaaggacaa gtcccaggga gcccaggcca ggcgggggcc 1440
cagaggggct tcccgggatg gtgggcgagg ccggggccgg ggccgcccag ggaagagagt 1500
ggcctgagcc caagccgcac cggagcagcg gctggattga acgccccaga ttggggcccg 1560
agatgtggcc ctcggtttcc tttcataaag gagttgtgtc cccagccctt ccactccagt 1620
aaagaactga attggccagg ggtccacgtc agcgtttggg atgggggatt ctggagccat 1680
acaaagcaac ccagagagtc ctgggccggc cacacccgag agtccctccc acctggtttc 1740
ttcctggaag ctgggtctct cccctaccct gcacggggtt ggtttcattg gtggcagcag 1800
cagccatgag tggccctccc cccagtcc 1828

<210> 33
<211> 2602
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3825977CB1

<400> 33
ccctcgagcg tgaaatggga aaggaatcat gcgactctgt aataagaagc tgacagaacc 60
atgaaggatc agggactgaa agcattagat aagttgttag tgggatgcag gttccacagg 120
caagcggaac gattcaagcg aattagtaaa gtttctgccc ggatcttgaa agccgcttcc 180
gttgctcagc ggaagtgtcg gtcgcaagag gacagacgcc tcgaagaatc cgctatcggc 240
tgtctgcaca accggaatca tgtcgagttt ggcggtgaga gacccggcaa tggatcgatc 300
actgcgttcc gtgttcgtgg ggaacattcc atatgaggca actgaggagc agttaaagga 360
cattttctcg gaggttggtt ctgttgtcag tttccggctg gtatacgata gagagacggg 420
aaaacccaag ggctatggct tctgcgaata ccaagaccag gagaccgcgc ttagtgccat 480
gcggaacctc aatgggcggg agttcagtgg gagagcgctt cgggtggaca atgctgccag 540
tgaaaagaat aaggaggagt taaagagcct tgggcctgca gcgcccatta ttgactcacc 600
ctatggggat cccatcgatc cagaagatgc ccctgaatcg attaccagag cagtagccag 660
tctcccccg gagcagatgt ttgagctgat gaagcagatg aagctctgtg tccaaaacag 720
ccaccaggaa gctcgaaaca tgttacttca aaatccacaa ctggcttatg cactgttgca 780
ggcacaagta gtgatgagaa tcatggatcc agagattgct ctgaaaattc tgcatcggaa 840
gatacatgtc acaccactga tcccaggcaa atctcagtct gtgtctgtct ctggccctgg 900
ccctggccct ggccctgggc tctgcccagg acctaatgtt ctgctgaacc agcagaatcc 960
tccagctcct cagcctcagc atttggctag aagacctgtg aaggacattc ctcctctgat 1020
gcagactcct atccagggtg gaattccagc tccagggcca ataccagctg cagttcccgg 1080
agctggtcct ggttccttaa ctcctggagg agcaatgcag ccccaacttg gaatgccagg 1140
ggttggccca gtgcctttag agcggggaca agtgcagatg tcagatccta gagctcctat 1200
acctcgcgga cccgtgactc ctggtggtct gcctcctcga ggactgttag gagatgctcc 1260
aaatgaccca cgtggaggga ctttgctttc agtcactgga gaagtggagc ccagaggtta 1320
tctgggtcca ccccatcagg gtccccccat gcatcatgcc tctggtcatg acactcgtgg 1380
cccttcctca catgagatga ggggagggcc attaggagat cccagactgc taattggaga 1440
gcccagaggc cccatgatag atcaaagggg tctacctatg gatggtagag gtggtagaga 1500
ttctcgagcg atggagactc gtgccatgga aactgaggtc ttagagacac gtgtaatgga 1560
gaggagagga atggagacct gtgcgatgga aaccagaggg atggaagcaa ggggcatgga 1620
tgcaagagga ttggagatga ggggccctgt ccccagttca agaggcccta tgactggtgg 1680
aattcagggt cctggtccca ttaatatagg ggcaggtggc cctcctcagg gacccagaca 1740
ggtcccaggc atttcagggg tggggaatcc tggagctggt atgcagggta caggcataca 1800

aggaacaggc atgcagggag caggcataca aggaggaggg atgcaggggg caggcataca 1860
aggagtcagt atacaaggag gaggtataca aggaggaggt atacaggggg caagcaagca 1920
aggtggaagc cagcctagca gttttagtcc tgggcagagc caggtcactc cacaggatca 1980
ggagaaggca gctttgatca tgcaggttct tcaactgact gcagatcaga ttgccatgct 2040
gccccctgag caaaggcaga gtatcctgat tttaaaggaa caaatccaga aatccactgg 2100
agcgtcttga aaggttttag aaaatatttg gctgtagtct caaattttat tctgtagcat 2160
ggagaatggg tgcaaaaagc tgacttctgt atccccacac ttggattagg gtttccctcc 2220
tcctagaacc taatcttatt ttttgttctt tttctttctt tctgttttcc tttttttttt 2280
aattgagggt gggggaggga gggagtgcgt ctgttcactt taagttactt taaaataact 2340
ctgaacatga ttatattatg ccaaataaga ttacaaagaa taagcagcaa tattgaagca 2400
tctacagtat gttaactaca ttttttaaat gtcgagtaaa acttcgtgaa aactgctcat 2460
aaagactaaa agttgacctg ttaaaacgtt aatgtactaa gatagtttta agatttttgg 2520
ttgtataaca aaataaaagt ttacccaaaa gtaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2580
aaaaaaaaaa aaaaaaaaaa aa 2602

<210> 34
<211> 566
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 4941262CB1

<400> 34
gcagaagtgg aaggaagcag gcggccacca tggtgaaggg acgcacggga cagcgcgtca 60
ggctctacgt ccggggcacc atcctcggat acaagaggtc caagtcgaac cagtacgaga 120
ccacgtcgct cattcagatc gagggggtga acaccaagga ggacgtcgcg tggtacgctg 180
gcaagcgcat ggcgtacatc tacaaggcta agaccaagag cagcgagacc cgctacaggt 240
gcatctgggg caaggtcacc cgcccgcacg gcaactcggg cgtcgtccgc gccaagttca 300
agtccaacct cccgcctgag tccatggggc gcaaggtcag agtgttcatg tacccgagca 360
gcatctaagg tttttgttgg agtaaaggtg gactctaaat ggccatgctt agttcttctc 420
tctgagctta aaatgccatg tgttggcaac ttagattgtt catgtactga acctgttgaa 480
gttctaccaa aatttgttgt cgaacggctg aacagttgtc ctaatgttat gctataaaca 540
gagcttattt caaaaaaaaa aaaaaa 566

<210> 35
<211> 183
<212> PRT
<213> Homo sapiens

<300>
<308> Incyte ID No.: g2961149

<400> 35
Met Ser Arg Tyr Leu Arg Pro Pro Asn Thr Ser Leu Phe Val Arg
1 5 10 15
Asn Val Ala Asp Asp Thr Arg Ser Glu Asp Leu Arg Arg Glu Phe
20 25 30
Gly Arg Tyr Gly Pro Ile Val Asp Val Tyr Val Pro Leu Asp Phe
35 40 45
Tyr Thr Arg Arg Pro Arg Gly Phe Ala Tyr Val Gln Phe Glu Asp
50 55 60
Val Arg Asp Ala Glu Asp Ala Leu His Asn Leu Asp Arg Lys Trp

```
                    65                  70                  75
Ile Cys Gly Arg Gln Ile Glu Ile Gln Phe Ala Gln Gly Asp Arg
                    80                  85                  90
Lys Thr Pro Asn Gln Met Lys Ala Lys Glu Gly Arg Asn Val Tyr
                    95                  100                 105
Ser Ser Ser Arg Tyr Asp Asp Tyr Asp Arg Tyr Arg Arg Ser Arg
                    110                 115                 120
Ser Arg Ser Tyr Glu Arg Arg Arg Ser Arg Ser Arg Ser Phe Asp
                    125                 130                 135
Tyr Asn Tyr Arg Arg Ser Tyr Ser Pro Arg Asn Ser Arg Pro Thr
                    140                 145                 150
Gly Arg Pro Arg Arg Ser Arg Ser His Ser Asp Asn Asp Arg Pro
                    155                 160                 165
Asn Cys Ser Trp Asn Thr Gln Tyr Ser Ser Ala Tyr Tyr Thr Ser
                    170                 175                 180
Arg Lys Ile


<210> 36
<211> 1404
<212> PRT
<213> Homo sapiens

<300>
<308> Incyte ID No.: g2660712

<400> 36
Met Ser Gly Ala Arg Thr Ala Ser Thr Pro Thr Pro Pro Gln Thr
  1             5                   10                  15
Gly Gly Gly Leu Glu Pro Gln Ala Asn Gly Glu Thr Pro Gln Val
                    20                  25                  30
Ala Val Ile Val Arg Pro Asp Asp Arg Ser Gln Gly Ala Ile Ile
                    35                  40                  45
Ala Asp Arg Pro Gly Leu Pro Gly Pro Glu His Ser Pro Ser Glu
                    50                  55                  60
Ser Gln Pro Ser Ser Pro Ser Pro Thr Pro Ser Pro Ser Pro Val
                    65                  70                  75
Leu Glu Pro Gly Ser Glu Pro Asn Leu Ala Val Leu Ser Ile Pro
                    80                  85                  90
Gly Asp Thr Met Thr Thr Ile Gln Met Ser Val Glu Glu Ser Thr
                    95                  100                 105
Pro Ile Ser Arg Glu Thr Gly Glu Pro Tyr Arg Leu Ser Pro Glu
                    110                 115                 120
Pro Thr Pro Leu Ala Glu Pro Ile Leu Glu Val Glu Val Thr Leu
                    125                 130                 135
Ser Lys Pro Val Pro Glu Ser Glu Phe Ser Ser Ser Pro Leu Gln
                    140                 145                 150
Ala Pro Thr Pro Leu Ala Ser His Thr Val Glu Ile His Glu Pro
                    155                 160                 165
Asn Gly Met Val Pro Ser Glu Asp Leu Glu Pro Glu Val Glu Ser
                    170                 175                 180
Ser Pro Glu Leu Ala Pro Pro Pro Ala Cys Pro Ser Glu Ser Pro
                    185                 190                 195
Val Pro Ile Ala Pro Thr Ala Gln Pro Glu Glu Leu Leu Asn Gly
                    200                 205                 210
Ala Pro Ser Pro Pro Ala Val Asp Leu Ser Pro Val Ser Glu Pro
                    215                 220                 225
```

Glu Glu Gln Ala Lys Glu Val Thr Ala Ser Val Ala Pro Pro Thr
230 235 240
Ile Pro Ser Ala Thr Pro Ala Thr Ala Pro Ser Ala Thr Ser Pro
245 250 255
Ala Gln Glu Glu Glu Met Glu Glu Glu Glu Glu Glu Glu Glu Gly
260 265 270
Glu Ala Gly Glu Ala Gly Glu Ala Glu Ser Glu Lys Gly Gly Glu
275 280 285
Glu Leu Leu Pro Pro Glu Ser Thr Pro Ile Pro Ala Asn Leu Ser
290 295 300
Gln Asn Leu Glu Ala Ala Ala Ala Thr Gln Val Ala Val Ser Val
305 310 315
Pro Lys Arg Arg Arg Lys Ile Lys Glu Leu Asn Lys Lys Glu Ala
320 325 330
Val Gly Asp Leu Leu Asp Ala Phe Lys Glu Ala Asn Pro Ala Val
335 340 345
Pro Glu Val Glu Asn Gln Pro Pro Ala Gly Ser Asn Pro Gly Pro
350 355 360
Glu Ser Glu Gly Ser Gly Val Pro Pro Arg Pro Glu Glu Ala Asp
365 370 375
Glu Thr Trp Asp Ser Lys Glu Asp Lys Ile His Asn Ala Glu Asn
380 385 390
Ile Gln Pro Gly Glu Gln Lys Tyr Glu Tyr Lys Ser Asp Gln Trp
395 400 405
Lys Pro Pro Asn Leu Glu Glu Lys Lys Arg Tyr Asp Arg Glu Phe
410 415 420
Leu Leu Gly Phe Gln Phe Ile Phe Ala Ser Met Gln Lys Pro Glu
425 430 435
Gly Leu Pro His Ile Ser Asp Val Val Leu Asp Lys Ala Asn Lys
440 445 450
Thr Pro Leu Arg Pro Leu Asp Pro Thr Arg Leu Gln Gly Ile Asn
455 460 465
Cys Gly Pro Asp Phe Thr Pro Ser Phe Ala Asn Leu Gly Arg Thr
470 475 480
Thr Leu Ser Thr Arg Gly Pro Pro Arg Gly Gly Pro Gly Gly Glu
485 490 495
Leu Pro Arg Gly Pro Gln Ala Gly Leu Gly Pro Arg Arg Ser Gln
500 505 510
Gln Gly Pro Arg Lys Glu Pro Arg Lys Ile Ile Ala Thr Val Leu
515 520 525
Met Thr Glu Asp Ile Lys Leu Asn Lys Ala Glu Lys Ala Trp Lys
530 535 540
Pro Ser Ser Lys Arg Thr Ala Ala Asp Lys Asp Arg Gly Glu Glu
545 550 555
Asp Ala Asp Gly Ser Lys Thr Gln Asp Leu Phe Arg Arg Val Arg
560 565 570
Ser Ile Leu Asn Lys Leu Thr Pro Gln Met Phe Gln Gln Leu Met
575 580 585
Lys Gln Val Thr Gln Leu Ala Ile Asp Thr Glu Glu Arg Leu Lys
590 595 600
Gly Val Ile Asp Leu Ile Phe Glu Lys Ala Ile Ser Glu Pro Asn
605 610 615
Phe Ser Val Ala Tyr Ala Asn Met Cys Arg Cys Leu Met Ala Leu
620 625 630
Lys Val Pro Thr Thr Glu Lys Pro Thr Val Thr Val Asn Phe Arg
635 640 645
Lys Leu Leu Leu Asn Arg Cys Gln Lys Glu Phe Glu Lys Asp Lys

650 655 660
Asp Asp Asp Glu Val Phe Glu Lys Lys Gln Lys Glu Met Asp Glu
665 670 675
Ala Ala Thr Ala Glu Glu Arg Gly Arg Leu Lys Glu Glu Leu Glu
680 685 690
Glu Ala Arg Asp Ile Ala Arg Arg Arg Ser Leu Gly Asn Ile Lys
695 700 705
Phe Ile Gly Glu Leu Phe Lys Leu Lys Met Leu Thr Glu Ala Ile
710 715 720
Met His Asp Cys Val Val Lys Leu Leu Lys Asn His Asp Glu Glu
725 730 735
Ser Leu Glu Cys Leu Cys Arg Leu Leu Thr Thr Ile Gly Lys Asp
740 745 750
Leu Asp Phe Glu Lys Ala Lys Pro Arg Met Asp Gln Tyr Phe Asn
755 760 765
Gln Met Glu Lys Ile Ile Lys Glu Lys Lys Thr Ser Ser Arg Ile
770 775 780
Arg Phe Met Leu Gln Asp Val Leu Asp Leu Arg Gly Ser Asn Trp
785 790 795
Val Pro Arg Arg Gly Asp Gln Gly Pro Lys Thr Ile Asp Gln Ile
800 805 810
His Lys Glu Ala Glu Met Glu Glu His Arg Glu His Ile Lys Val
815 820 825
Gln Gln Leu Met Ala Lys Gly Ser Asp Lys Arg Arg Gly Gly Pro
830 835 840
Pro Gly Pro Pro Ile Ser Arg Gly Leu Pro Leu Val Asp Asp Gly
845 850 855
Gly Trp Asn Thr Val Pro Ile Ser Lys Gly Ser Arg Pro Ile Asp
860 865 870
Thr Ser Arg Leu Thr Lys Ile Thr Lys Pro Gly Ser Ile Asp Ser
875 880 885
Asn Asn Gln Leu Phe Ala Pro Gly Gly Arg Leu Ser Trp Gly Lys
890 895 900
Gly Ser Ser Gly Gly Ser Gly Ala Lys Pro Ser Asp Ala Ala Ser
905 910 915
Glu Ala Ala Arg Pro Ala Thr Ser Thr Leu Asn Arg Phe Ser Ala
920 925 930
Leu Gln Gln Ala Val Pro Thr Glu Ser Thr Asp Asn Arg Arg Val
935 940 945
Val Gln Arg Ser Ser Leu Ser Arg Glu Arg Gly Glu Lys Ala Gly
950 955 960
Asp Arg Gly Asp Arg Leu Glu Arg Ser Glu Arg Gly Gly Asp Arg
965 970 975
Gly Asp Arg Leu Asp Arg Ala Arg Thr Pro Ala Thr Lys Arg Ser
980 985 990
Phe Ser Lys Glu Val Glu Glu Arg Ser Arg Glu Arg Pro Ser Gln
995 1000 1005
Pro Glu Gly Leu Arg Lys Ala Ala Ser Leu Thr Glu Asp Arg Asp
1010 1015 1020
Arg Gly Arg Asp Ala Val Lys Arg Glu Ala Ala Leu Pro Pro Val
1025 1030 1035
Ser Pro Leu Lys Ala Ala Leu Ser Glu Glu Glu Leu Glu Lys Lys
1040 1045 1050
Ser Lys Ala Ile Ile Glu Glu Tyr Leu His Leu Asn Asp Met Lys
1055 1060 1065
Glu Ala Val Gln Cys Val Gln Glu Leu Ala Ser Pro Ser Leu Leu
1070 1075 1080

```
Phe Ile Phe Val Arg His Gly Val Glu Ser Thr Leu Glu Arg Ser
                1085                1090                1095
Ala Ile Ala Arg Glu His Met Gly Gln Leu Leu His Gln Leu Leu
                1100                1105                1110
Cys Ala Gly His Leu Ser Thr Ala Gln Tyr Tyr Gln Gly Leu Tyr
                1115                1120                1125
Glu Ile Leu Glu Leu Ala Glu Asp Met Glu Ile Asp Ile Pro His
                1130                1135                1140
Val Trp Leu Tyr Leu Ala Glu Leu Val Thr Pro Ile Leu Gln Glu
                1145                1150                1155
Gly Gly Val Pro Met Gly Glu Leu Phe Arg Glu Ile Thr Lys Pro
                1160                1165                1170
Leu Arg Pro Leu Gly Lys Ala Ala Ser Leu Leu Leu Glu Ile Leu
                1175                1180                1185
Gly Leu Leu Cys Lys Ser Met Gly Pro Lys Lys Val Gly Thr Leu
                1190                1195                1200
Trp Arg Glu Ala Gly Leu Ser Trp Lys Glu Phe Leu Pro Glu Gly
                1205                1210                1215
Gln Asp Ile Gly Ala Phe Val Ala Glu Gln Lys Val Glu Tyr Thr
                1220                1225                1230
Leu Gly Glu Glu Ser Glu Ala Pro Gly Gln Arg Ala Leu Pro Ser
                1235                1240                1245
Glu Glu Leu Asn Arg Gln Leu Glu Lys Leu Leu Lys Glu Gly Ser
                1250                1255                1260
Ser Asn Gln Arg Val Phe Asp Trp Ile Glu Ala Asn Leu Ser Glu
                1265                1270                1275
Gln Gln Ile Val Ser Asn Thr Leu Val Arg Ala Leu Met Thr Ala
                1280                1285                1290
Val Cys Tyr Ser Ala Ile Ile Phe Glu Thr Pro Leu Arg Val Asp
                1295                1300                1305
Val Ala Val Leu Lys Ala Arg Ala Lys Leu Leu Gln Lys Tyr Leu
                1310                1315                1320
Cys Asp Glu Gln Lys Glu Leu Gln Ala Leu Tyr Ala Leu Gln Ala
                1325                1330                1335
Leu Val Val Thr Leu Glu Gln Pro Pro Asn Leu Leu Arg Met Phe
                1340                1345                1350
Phe Asp Ala Leu Tyr Asp Glu Asp Val Val Lys Glu Asp Ala Phe
                1355                1360                1365
Tyr Ser Trp Glu Ser Ser Lys Asp Pro Ala Glu Gln Gln Gly Lys
                1370                1375                1380
Gly Val Ala Leu Lys Ser Val Thr Ala Phe Phe Lys Trp Leu Arg
                1385                1390                1395
Glu Ala Glu Glu Glu Ser Asp His Asn
                1400
```

<210> 37
<211> 322
<212> PRT
<213> Homo sapiens

<300>
<308> Incyte ID No.: g2440051

<400> 37
Leu Arg Arg Gly Ala Arg Asn Trp Arg Ser Met Ser Thr Gly Glu
1 5 10 15
Leu Thr Pro Gln Ser Arg Leu Lys Glu Phe Ser Glu Leu Ala Arg
20 25 30
Ala Leu Asn Leu Tyr Arg Met Asp His Leu Gly Asn Tyr Thr Gly
35 40 45
His Lys Ser Tyr Tyr Leu Thr Gly Gln Leu Ala Thr Leu Glu Gln
50 55 60
Ala Ile Ile Gln Tyr Ala Leu Gln Ala Val Thr Glu His Gly Phe
65 70 75
Lys Leu Ile Ser Val Pro Asp Ile Leu Pro Lys Glu Val Ile Glu
80 85 90
Ser Cys Gly Met Arg Thr Glu Gly Glu Arg Thr Gln Val Tyr Lys
95 100 105
Leu Asp Thr Gly Glu Cys Leu Ser Gly Thr Ser Glu Met Ala Leu
110 115 120
Ala Gly Phe Phe Ala Asn Lys Leu Leu Ser Glu Asp Gln Leu Pro
125 130 135
Leu Lys Val Thr Ala Val Ser Arg Cys Tyr Arg Ala Glu Thr Ser
140 145 150
Gly Leu Gln Glu Glu Lys Gly Ile Tyr Arg Val His Gln Phe Asn
155 160 165
Lys Val Glu Met Phe Ala Ile Cys Thr Glu Glu Gln Ser Glu Ala
170 175 180
Glu Leu Glu Glu Phe Lys Asn Ile Glu Val Asp Leu Phe Arg Arg
185 190 195
Leu Gly Leu Asn Phe Arg Leu Leu Asp Met Pro Pro Cys Glu Leu
200 205 210
Gly Ala Pro Ala Tyr Gln Lys Tyr Asp Ile Glu Ala Trp Met Pro
215 220 225
Gly Arg Gln Met Trp Gly Glu Ile Ser Ser Cys Ser Asn Cys Thr
230 235 240
Asp Tyr Gln Ala Arg Arg Leu Gly Ile Arg Tyr Arg Arg Ser Ala
245 250 255
Asp Gly Gln Ile Leu His Ala His Thr Ile Asn Gly Thr Ala Thr
260 265 270
Ala Ile Pro Arg Leu Leu Ile Ala Leu Leu Glu Ser Tyr Gln Lys
275 280 285
Glu Asp Gly Ile Glu Ile Pro Ala Val Leu Arg Pro Phe Met Asp
290 295 300
Asn Gln Glu Leu Ile Thr Arg Asn Lys Arg Ile Pro Glu Thr Lys
305 310 315
Leu Val Lys Phe Ile Lys Ala
320

<210> 38
<211> 343
<212> PRT
<213> Homo sapiens

<300>
<308> Incyte ID No.: g1808648

<400> 38

Met Glu Val Ser Cys Gly Gln Ala Glu Ser Ser Glu Lys Pro Asn
1 5 10 15

Ala Glu Asp Met Thr Ser Lys Asp Tyr Tyr Phe Asp Ser Tyr Ala
20 25 30

His Phe Gly Ile His Glu Glu Met Leu Lys Asp Glu Val Arg Thr
35 40 45

Leu Thr Tyr Arg Asn Ser Met Phe His Asn Arg His Leu Phe Lys
50 55 60

Asp Lys Val Val Leu Asp Val Gly Ser Gly Thr Gly Ile Leu Cys
65 70 75

Met Phe Ala Ala Lys Ala Gly Ala Arg Lys Val Ile Gly Ile Val
80 85 90

Cys Ser Ser Ile Ser Asp Tyr Ala Val Lys Ile Val Lys Ala Asn
95 100 105

Lys Leu Asp His Val Val Thr Ile Ile Lys Gly Lys Val Glu Glu
110 115 120

Val Glu Leu Pro Val Glu Lys Val Asp Ile Ile Ile Ser Glu Trp
125 130 135

Met Gly Tyr Cys Leu Phe Tyr Glu Ser Met Leu Asn Thr Val Leu
140 145 150

Tyr Ala Arg Asp Lys Trp Leu Ala Pro Asp Gly Leu Ile Phe Pro
155 160 165

Asp Arg Ala Thr Leu Tyr Val Thr Ala Ile Glu Asp Arg Gln Tyr
170 175 180

Lys Asp Tyr Lys Ile His Trp Trp Glu Asn Val Tyr Gly Phe Asp
185 190 195

Met Ser Cys Ile Lys Asp Val Ala Ile Lys Glu Pro Leu Val Asp
200 205 210

Val Val Asp Pro Lys Gln Leu Val Thr Asn Ala Cys Leu Ile Lys
215 220 225

Glu Val Asp Ile Tyr Thr Val Lys Val Glu Asp Leu Thr Phe Thr
230 235 240

Ser Pro Phe Cys Leu Gln Val Lys Arg Asn Asp Tyr Val His Ala
245 250 255

Leu Val Ala Tyr Phe Asn Ile Glu Phe Thr Arg Cys His Lys Arg
260 265 270

Thr Gly Phe Ser Thr Ser Pro Glu Ser Pro Tyr Thr His Trp Lys
275 280 285

Gln Thr Val Phe Tyr Met Glu Asp Tyr Leu Thr Val Lys Thr Gly
290 295 300

Glu Glu Ile Phe Gly Thr Ile Gly Met Arg Pro Asn Ala Lys Asn
305 310 315

Asn Arg Asp Leu Asp Phe Thr Ile Asp Leu Asp Phe Lys Gly Gln
320 325 330

Leu Cys Glu Leu Ser Cys Ser Thr Asp Tyr Arg Met Arg
335 340